

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: January 29, 2002, 10:21:42 ; Search time 310.82 seconds

(without alignments)  
5.481 Million cell updates/sec

Title: US-09-763-397A-6

Perfect score: 127  
Sequence: 1 KPKDELVDENDIKICKMEKCS 23

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq\_1101.\*

1: /SID8/gcgdata/geneseq/genesep/AA1980.DAT.\*  
2: /SID8/gcgdata/geneseq/genesep/AA1981.DAT.\*  
3: /SID8/gcgdata/geneseq/genesep/AA1982.DAT.\*  
4: /SID8/gcgdata/geneseq/genesep/AA1983.DAT.\*  
5: /SID8/gcgdata/geneseq/genesep/AA1984.DAT.\*  
6: /SID8/gcgdata/geneseq/genesep/AA1985.DAT.\*  
7: /SID8/gcgdata/geneseq/genesep/AA1986.DAT.\*  
8: /SID8/gcgdata/geneseq/genesep/AA1987.DAT.\*  
9: /SID8/gcgdata/geneseq/genesep/AA1988.DAT.\*  
10: /SID8/gcgdata/geneseq/genesep/AA1989.DAT.\*  
11: /SID8/gcgdata/geneseq/genesep/AA1990.DAT.\*  
12: /SID8/gcgdata/geneseq/genesep/AA1991.DAT.\*  
13: /SID8/gcgdata/geneseq/genesep/AA1992.DAT.\*  
14: /SID8/gcgdata/geneseq/genesep/AA1993.DAT.\*  
15: /SID8/gcgdata/geneseq/genesep/AA1994.DAT.\*  
16: /SID8/gcgdata/geneseq/genesep/AA1995.DAT.\*  
17: /SID8/gcgdata/geneseq/genesep/AA1996.DAT.\*  
18: /SID8/gcgdata/geneseq/genesep/AA1997.DAT.\*  
19: /SID8/gcgdata/geneseq/genesep/AA1998.DAT.\*  
20: /SID8/gcgdata/geneseq/genesep/AA1999.DAT.\*  
21: /SID8/gcgdata/geneseq/genesep/AA2000.DAT.\*  
22: /SID8/gcgdata/geneseq/genesep/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	127	100.0	23	AAV70282
2	127	100.0	66	AA66442
3	127	100.0	309	AA13175
4	127	100.0	319	AA13176
5	127	100.0	319	AA13177
6	127	100.0	327	AA13178
7	127	100.0	335	AA13179
8	127	100.0	335	AA13179
9	127	100.0	350	AAV70278
10	127	100.0	411	AA68314
11	127	100.0	412	AA60416

12	127	100.0	424	AA37796
13	123	96.9	412	AA80835
14	121	95.3	424	AA37797
15	120	94.5	180	AA07290
16	96	75.6	40	AA87213
17	75	59.1	20	AA71651
18	71	55.9	21	AAV70283
19	62	48.8	14	AA71655
20	62	48.8	19	AA61002
21	60.5	47.6	402	AA70709
22	58.5	46.1	19	AA71654
23	58.5	46.1	168	AA07291
24	58.5	46.1	272	AA93560
25	56.5	44.5	478	AA07288
26	53	41.7	19	AA99033
27	53	41.7	20	AAJ0418
28	53	41.7	21	AA91504
29	53	41.7	21	AA82586
30	53	41.7	21	AA87820
31	53	41.7	21	AA87595
32	53	41.7	21	AA870912
33	53	41.7	21	AA805612
34	53	41.7	21	AA835440
35	53	41.7	21	AAV23252
36	53	41.7	21	AA80071
37	53	41.7	21	AAV54553
38	53	41.7	21	AAV58777
39	53	41.7	21	AA899706
40	53	41.7	21	AA62428
41	53	41.7	21	AA684517
42	53	41.7	21	AA688269
43	53	41.7	21	AA689366
44	53	41.7	21	AA84447
45	53	41.7	21	AA898457

#### ALIGNMENTS

RESULT 1	AAV70282	standard; peptide; 23 AA
ID	AAV70282	
XX	AAV70282	
AC	AAV70282	
XX	AAV70282	
DT	06-JUN-2000	(first entry)
XX	AAV70282	
DE	Plasmodium falciparum CSP antigenic epitope, p593.	
XX	Recombinant protein: CDC/NITMVA-C-1; multivalent; malaria; vaccine;	
KW	T-cell epitope; tetanus toxoid; antigenic epitope; treatment;	
KW	Circumsporozoite protein; CSP; sporozoite surface protein-2; SSP-2;	
KW	liver stage antigen-1; USA-1; merozoite surface protein-1; MSP-1;	
KW	apical membrane antigen-1; AMA-1; erythrocyte binding antigen-175;	
KW	EBA-175; rhoptry associated protein-1; RAP-1; gamete specific antigen;	
KW	pfg27; antiparasitic; prevention; anti-CDC/NITMVA-C-1 antibody.	
XX	Plasmodium falciparum.	
OS	WO20001179-A1.	
XX	02-MAR-2000.	
PD	19-AUG-1999.	99MO-US18869.
XX	21-AUG-1998.	98US-0097703.
PR	(NAIM-) NAT INST IMMUNOLOGY.	
XX	(USSH) US DEPT HEALTH & HUMAN SERVICES.	
PA	Lat AA, Sh1 YP, Hasnain SE;	
XX	WPI; 2000-237654/20.	

RIS protein. Syn

Sequence encoded b

RIS\* protein. Syn

Circumsporozoite a

P. falciparum deriv

CS Region II\* mme

Plasmodium falcipa

Circumsporozoite a

Sequence corresp.

Plasmodium cynomol

CS Region II\* mme

Circumsporozoite a

Plasmodium berghei

Human IL-2/ P. yoe

HuA class II bindi

P. falciparum CS pr

Sequence of modifi

Plasmodium falcipa

Malaria circumspor

P. falciparum CS p

Malaria circumspor

Circumsporozoite h

T-cell stimulatory

Peptide derived fr

Pathogen derived T

T helper cell (Th)

Unidentified pepti

Plasmodium falcipa

Plasmodium falcipa

Plasmodium falcipa

Plasmodium falcipa

Plasmodium falcipa

Sequence of T help

Plasmodium falcipa



XX Example 1; Page 7; 18pp; English.

PS The polypeptide is prep'd. by genetic engineering of genes encoding  
 CC the P. falciparum circumsporozoite (CS) protein [Deme et al.,  
 CC Science 225 : 593 (1984)], and the influenza virus non-structural  
 CC protein 1 (NS1). [Baez et al., Nucleic Acids Research, 8 : 5845  
 CC (1980)]. The DNA encoding the 1st 81 AAs of the N-terminal of NS1  
 CC (NS1-81) is linked via a synthetic sequence to DNA encoding Region  
 CC I contg. flanking regionless the 18 AA signal region, which in  
 CC turn is fused to DNA encoding Region II contg. flanking region  
 CC less the first nine N-terminal AAs. This CS fusion is designated  
 CC Rlfelta9. The Pro residue separating the Asp at the C-terminal  
 CC of the linker from Rlfelta9 is an artifact of a filled in BamHI  
 CC site. The peptide can be used in a vaccine for protection against  
 CC malaria.  
 CC See also AAR12306-R12311 and AAR13176-R13179.

CC Sequence 309 AA;

SO

Query Match 100.0%; Score 127; DB 12; Length 309;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-10;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPKDELVDYNDIEKKICKMEKCS 23  
 |||  
 DB 265 kpkdelvyndiekkickmekcs 287

RESULT 4  
 AAR07945  
 ID AAR07945 standard; protein: 319 AA.

AC AAR07945;  
 XX  
 DT 22-FEB-1991 (first entry)  
 XX  
 DE NS181RLFauth plasmid product.  
 XX  
 KW Malaria; vaccine.  
 XX  
 OS Plasmodium falciparum.  
 XX

FT Key Location/Qualifiers  
 FT Domain 1..81  
 FT /label= NS181 protein fragment  
 FT /note= "from plasmid pmg-1"  
 FT Domain 89..193  
 FT /label= Fragment of circumsporozoite protein  
 FT Domain 204..319  
 FT /label= Fragment of circumsporozoite protein

EP398540-A.  
 XX  
 PD 22-NOV-1990.  
 XX  
 PF 01-MAY-1990; 90EP-0304720.  
 XX  
 PR 03-MAY-1989; 89US-0346863.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM.  
 XX  
 PI Gross MS, Young JF;  
 XX  
 DR WPI; 1990-350299/47.  
 XX  
 DR N-PSDB; AA006580.  
 XX

PT New polypeptide used in malaria vaccine - comprises immunogenic  
 PT determinants from 1st and 2nd flanking regions of plasmodium  
 PT surface protein and intermediate repeat domain  
 XX  
 XX Example 2; Page 11-12; 24pp; English.

XX The product is useful in preparation of vaccines for treatment and  
 CC prophylaxis of plasmodium sporozite infection. It may be easily  
 CC produced in large pure quantities from a transformed E.coli  
 CC expression system.  
 XX

SO Sequence 319 AA;

Query Match 100.0%; Score 127; DB 11; Length 319;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-10;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPKDELVDYNDIEKKICKMEKCS 23  
 |||  
 DB 275 kpkdelvyndiekkickmekcs 297

RESULT 5  
 AAR13176  
 ID AAR13176 standard; protein: 319 AA.

AC AAR13176;  
 XX  
 DT 29-AUG-1991 (first entry)  
 XX  
 DE NS1\_81-RLFauth.  
 XX  
 KW Immunogenic determinant; circumsporozoite; CS; vaccine; malaria;  
 KW hybrid; Influenza virus; non-structural protein 1; fusion.  
 XX  
 OS Plasmodium falciparum.  
 OS Influenza virus (A/PR/8/34);

FT Key Location/Qualifiers  
 FT Region 1..81  
 FT /label= N-terminal of NS1  
 FT /note= "Influenza virus nonstructural protein 1"  
 FT Peptide 82..87  
 FT /label= synthetic linker  
 FT Region 88  
 FT /label= artifact  
 FT /note= "see comments"  
 FT Region 89..193  
 FT /label= AAs 19-123 of CS protein  
 FT /note= "Region I contg. flanking region less  
 FT signal sequence"  
 FT Region 194  
 FT /label= artifact  
 FT /note= "see comments"  
 FT Region 195..319  
 FT /label= AAs 288-412 of CS protein  
 FT /note= "Region II flanking region"

EP432965-A.  
 XX  
 PD 19-JUN-1991.  
 XX  
 PF 06-DEC-1990; 90EP-0313257.  
 XX  
 PR 08-DEC-1989; 89US-0447746.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM.  
 PA (USSA ) US SEC OF THE ARMY.  
 PA (BIOM-) BIOMEDICAL RES INST.  
 XX  
 PI Gross MS, Gordon DM, Hollingdale MR;  
 XX  
 DR WPI; 1991-179771/25.  
 XX  
 DR Polypeptide comprising immunogenic determinants from P falciparum  
 PT - for vaccine against malaria infection in humans.  
 XX

PS Example 2; Page 10; 18pp; English.

XX CC The polypeptide is prep'd. by genetic engineering of genes encoding  
 CC the P. falciparum circumsporozoite (CS) protein [Dame et al.,  
 CC Science 225 : 593 (1984)], and the influenza virus non-structural  
 CC protein 1 (NS1), [Baez et al., Nucleic Acids Research, 8 : 5845  
 CC (1980)]. The DNA encoding the 1st 81 AAs of the N-terminal of NS1  
 CC (NS1\_81) is linked via a synthetic sequence to DNA encoding Region  
 CC I contg. flanking regions less the 18 AA signal region, which in  
 CC turn is fused to DNA encoding Region II-contg. flanking region.  
 CC This CS fusion is designated R1Aauth. The Pro residue separating  
 CC the Asp (at the C-terminal of the linker) from R1Aauth is an arti-  
 CC fact of a filled in BamHI site; the Gly separating Region I and  
 CC Region II-contg. CS flanking regions is an artifact of a synthetic  
 CC FokI/NotI linker. The peptide can be used in a vaccine for  
 CC protection against malaria.

CC The complete nucleotide and AA sequences are given in EP-304720,  
 CC filed May 1, 1990.  
 CC See also AAR12306-R12311 and AAR13175-R13179.

XX SQ Sequence 319 AA;

OY 1 KPKELDYENDIEKIKCMKCS 23  
 DB 275 kpkdelgyendiekkickmekcs 297

Query Match 100.0%; Score 127; DB 12; Length 319;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-10;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 6  
 AAR13177  
 ID AAR13177 standard; Protein; 327 AA.

XX AC AAR13177;  
 XX DT 29-AUG-1991 (first entry)

XX DE NS1\_81-R1Aauth + (NANP)2.  
 XX KW Immunogenic determinant; circumsporozoite; CS; vaccine; malaria;  
 XX hybrid; influenza virus; non-structural protein 1; fusion.  
 XX OS Plasmodium falciparum.  
 XX influenza virus (A/PR/8/34/).

XX FH Key Location/Qualifiers  
 XX Region 1..81  
 FT /label= "N-terminal of NS1  
 FT /note= "Influenza virus nonstructural protein 1"  
 FT Peptide 82..87  
 FT /label= synthetic linker  
 FT Region 88  
 FT /label= artifact  
 FT /note= "see comments"  
 FT Region 89..193  
 FT /label= AAs 19-123 of CS protein  
 FT /note= "Region I contg. flanking region less  
 FT signal sequence"  
 FT Region 194..201  
 FT /label= immunodominant repeat region  
 FT /note= "two tetrapeptide repeat units"  
 FT Region 202  
 FT /label= artifact  
 FT /note= "see comments"  
 FT Region 203..327  
 FT /label= AAs 288-412 of CS protein  
 FT /note= "Region II flanking region"  
 XX EP432965-A.  
 XX PN

PD 19-JUN-1991.  
 XX  
 XX PF 06-DEC-1990; 90EP-0313257.  
 XX PR 08-DEC-1989; 89US-0447746.  
 XX XX (SMIR ) SMITHKLINE BEECHAM.  
 XX PA (USSA ) US SEC OF THE ARMY.  
 XX PA (BIOM-) BIOMEDICAL RES INST.  
 XX PI Gross MS, Gordon DM, Hollingdale MR;  
 XX WPI; 1991-179771/25.  
 XX DR  
 XX PT Polypeptide comprising immunogenic determinants from P falciparum  
 XX PT - for vaccine against malaria infection in humans.  
 XX PS Example 3; Page 10; 18pp; English.

XX CC The polypeptide is prep'd. by genetic engineering of genes encoding  
 CC the P. falciparum circumsporozoite (CS) protein [Dame et al.,  
 CC Science 225 : 593 (1984)], and the influenza virus non-structural  
 CC protein 1 (NS1), [Baez et al., Nucleic Acids Research, 8 : 5845  
 CC (1980)]. The DNA encoding the 1st 81 AAs of the N-terminal of NS1  
 CC (NS1\_81) is linked via a synthetic sequence to DNA encoding Region  
 CC I contg. flanking region less the 18 AA signal region. This is  
 CC linked to a synthetic sequence encoding two repeat units from the  
 CC immunodominant region, which in turn is fused to DNA encoding the  
 CC Region II-contg. flanking region. The Pro residue separating the  
 CC Asp (at the C-terminal of the linker) from the Region I-contg. CS  
 CC flanking region is an artifact of a filled-in BamHI site; the Gly  
 CC separating the repeat units and the Region II-contg. CS flanking  
 CC region is an artifact of a synthetic FokI/NotI linker. The  
 CC peptide can be used in a vaccine for protection against malaria.  
 CC See also AAR12306-R12311 and AAR13175-R13179.

XX SQ Sequence 327 AA;

OY 1 KPKELDYENDIEKIKCMKCS 23  
 DB 283 kpkdelgyendiekkickmekcs 305

Query Match 100.0%; Score 127; DB 12; Length 327;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-10;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 7  
 AAR13178  
 ID AAR13178 standard; Protein; 335 AA.

XX AC AAR13178;  
 XX DT 29-AUG-1991 (first entry)

XX DE NS1\_81(NANP)4R1Aauth.  
 XX KW Immunogenic determinant; circumsporozoite; CS; vaccine; malaria;  
 XX hybrid; influenza virus; non-structural protein 1; fusion.  
 XX OS Plasmodium falciparum.  
 XX influenza virus (A/PR/8/34/).

XX FH Key Location/Qualifiers  
 XX Region 1..81  
 FT /label= "N-terminal of NS1  
 FT /note= "Influenza virus nonstructural protein 1"  
 FT Region 82..97  
 FT /label= immunodominant repeat region  
 FT /note= "four tetrapeptide repeat units"  
 FT Peptide 98..103  
 FT /label= synthetic linker



RA Olafsson P., Matile H., Certa U.;  
RT Plasmodium falciparum: the repetitive MSA-1 surface protein of the  
RT RO-71 isolate is recognized by mouse antibody against the  
RT nonrepetitive repeat block of RO-33.";  
RL Exp. Parasitol. 74:381-389(1992).  
DR EMBL: X61930; CAA43932.1; -  
DR InterPro: IPR000087; Collagen.  
FT NON TER 1087  
SQ SEQUENCE 1087 AA; 123911 MW; 5518852133C01B33 CRC64;

Query Match 57.4%; Score 42.5; DB 5; Length 1087;  
Best Local Similarity 60.0%; Pred. No. 53;  
Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

QY 1 OYKANSKFI-GITE 14  
ID 09T2T5 PRELIMINARY; PRT; 1694 AA.  
AC 09T2T5;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR.  
GN MSP1.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HN1;  
RA Jiang G., Liu R.Z., Daubenberger C.A., Pluschke G.;  
RL Submitted (Apr-1998) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AF062348; AAC72884.1; -  
DR InterPro: IPR000561; EGF-like.  
DR Pfam: PF00008; EGF; 1.  
KW Merozoite.  
SQ SEQUENCE 1694 AA; 192794 MW; 84CFC0E709F5673B CRC64;

Query Match 57.4%; Score 42.5; DB 5; Length 1694;  
Best Local Similarity 60.0%; Pred. No. 84;  
Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

QY 1 OYKANSKFI-GITE 14  
ID 09NHX1 PRELIMINARY; PRT; 1694 AA.  
AC 09NHX1;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE MAJOR MEROZOITE SURFACE ANTIGEN.  
GN GPI95.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FCCL/HN;  
RA Shan Z.X., Yu X.B., Li X.R., Ma C.L., Fang J.M.;  
RT "Molecular cloning and sequence analysis of major merozoite surface  
RT antigen(gpi95)gene of Plasmodium falciparum isolate FCCL/HN.";  
RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AF218248; AAF37526.1; -

RESULT 14  
ID 09NHX1 PRELIMINARY; PRT; 1694 AA.  
AC 09NHX1;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE MAJOR MEROZOITE SURFACE ANTIGEN.  
GN GPI95.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FCCL/HN;  
RA Shan Z.X., Yu X.B., Li X.R., Ma C.L., Fang J.M.;  
RT "Molecular cloning and sequence analysis of major merozoite surface  
RT antigen(gpi95)gene of Plasmodium falciparum isolate FCCL/HN.";  
RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AF218248; AAF37526.1; -

DR InterPro: IPR000561; EGF-like.  
DR Pfam: PF00008; EGF; 1.  
KW Merozoite.  
SQ SEQUENCE 1694 AA; 192766 MW; B51634A49E0F6728 CRC64;

Query Match 57.4%; Score 42.5; DB 5; Length 1694;  
Best Local Similarity 60.0%; Pred. No. 84;  
Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

QY 1 OYKANSKFI-GITE 14  
ID 09T2T4 PRELIMINARY; PRT; 1704 AA.  
AC 09T2T4;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR.  
GN MSP1.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HN2;  
RA Jiang G., Liu R.Z., Daubenberger C.A., Pluschke G.;  
RL Submitted (Apr-1998) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AF062349; AAC72885.1; -  
DR InterPro: IPR000561; EGF-like.  
DR Pfam: PF00008; EGF; 1.  
KW Merozoite.  
SQ SEQUENCE 1704 AA; 193762 MW; 3855260DA56FD1D CRC64;

Query Match 57.4%; Score 42.5; DB 5; Length 1704;  
Best Local Similarity 60.0%; Pred. No. 85;  
Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

QY 1 OYKANSKFI-GITE 14  
ID 09T2T4 PRELIMINARY; PRT; 1704 AA.  
AC 09T2T4;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR.  
GN MSP1.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HN2;  
RA Jiang G., Liu R.Z., Daubenberger C.A., Pluschke G.;  
RL Submitted (Apr-1998) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AF062349; AAC72885.1; -  
DR InterPro: IPR000561; EGF-like.  
DR Pfam: PF00008; EGF; 1.  
KW Merozoite.  
SQ SEQUENCE 1704 AA; 193762 MW; 3855260DA56FD1D CRC64;

Search completed: January 29, 2002, 11:15:53  
Job time: 213 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 10:55:58 ; Search time 1760.55 Seconds  
(without alignments)  
2.523 Million cell updates/sec

Title: US-09-763-397A-3

Perfect score: 97  
Sequence: 1 KPLDKFGNIYDYHEH 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 3148936 seqs, 277657034 residues

Total number of hits satisfying chosen parameters: 3148936

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_Main: \*  
1: /cgn2\_6/ptodata/2/paa/PCRNUS\_COMB.pep: \*  
2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pep: \*  
3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pep: \*  
4: /cgn2\_6/ptodata/2/paa/US080\_COMB.pep: \*  
5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pep: \*  
6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pep: \*  
7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pep: \*  
8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pep: \*  
9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pep: \*  
10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep: \*  
11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pep: \*  
12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pep: \*  
13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep: \*  
14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep: \*  
15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pep: \*  
16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pep: \*  
17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pep: \*  
18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pep: \*  
19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pep: \*  
20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pep: \*  
21: /cgn2\_6/ptodata/2/paa/US097\_COMB.pep: \*  
22: /cgn2\_6/ptodata/2/paa/US098\_COMB.pep: \*  
23: /cgn2\_6/ptodata/2/paa/US099\_COMB.pep: \*  
24: /cgn2\_6/ptodata/2/paa/US60\_COMB.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	97	100.0	16	21	US-09-763-397A-3
2	97	100.0	350	21	US-09-763-397A-2
3	54	55.7	277	24	US-60-324-109-22535
4	54	55.7	281	24	US-60-312-544-9219
5	51	52.6	313	15	US-09-134-001C-3628
6	48	49.5	262	15	US-09-198-452C-646
7	48	49.5	262	18	US-09-438-185-608
8	47	48.5	78	17	US-09-328-352-7148
9	46	47.4	102	17	US-09-328-352-5557

10	46	47.4	134	17	US-09-328-352-5702	Sequence 5702, Ap
11	46	47.4	141	17	US-09-328-352-7241	Sequence 7241, Ap
12	46	47.4	146	17	US-09-328-352-4183	Sequence 4183, Ap
13	46	47.4	146	17	US-09-328-352-4906	Sequence 4906, Ap
14	46	47.4	191	17	US-09-328-352-7059	Sequence 7059, Ap
15	46	47.4	209	17	US-09-328-352-7440	Sequence 7440, Ap
16	46	47.4	222	17	US-09-328-352-6979	Sequence 6979, Ap
17	46	47.4	345	17	US-09-328-352-5480	Sequence 5480, Ap
18	46	47.4	345	17	US-09-328-352-7604	Sequence 7604, Ap
19	45.5	46.9	247	17	US-09-328-352-7311	Sequence 7311, Ap
20	45	46.4	512	19	US-09-570-581A-938	Sequence 938, App
21	44.5	45.9	534	1	PCT-US01-08631-44074	Sequence 44074, A
22	44	45.4	86	24	US-60-196-718-6204	Sequence 6204, Ap
23	44	45.4	89	22	US-09-856-404-20	Sequence 5215, Ap
24	44	45.4	138	24	US-60-215-161-5215	Sequence 10, Appl
25	44	45.4	309	22	US-09-856-404-10	Sequence 4086, Ap
26	44	45.4	459	19	US-09-583-110-4086	Sequence 4093, Ap
27	44	45.4	463	15	US-09-107-433-4093	Sequence 15941, A
28	44	45.4	1263	21	US-09-733-089-15941	Sequence 15941, A
29	44	45.4	1263	22	US-09-816-660-15941	Sequence 15941, A
30	44	45.4	1614	21	US-09-733-089-200	Sequence 200, App
31	44	45.4	1614	22	US-09-816-660-200	Sequence 200, App
32	43	44.3	77	24	US-60-170-373-2877	Sequence 2877, Ap
33	43	44.3	77	24	US-60-173-469-1167	Sequence 1167, Ap
34	43	44.3	77	24	US-60-188-162-3835	Sequence 3835, Ap
35	43	44.3	78	24	US-60-169-840-4940	Sequence 4940, Ap
36	43	44.3	199	15	US-09-107-532-5931	Sequence 5931, Ap
37	43	44.3	199	15	US-09-107-532A-5931	Sequence 5931, Ap
38	43	44.3	243	24	US-60-241-098-14	Sequence 14, Appl
39	43	44.3	329	1	PCT-US01-03404B-14	Sequence 14, Appl
40	43	44.3	3952	24	US-60-173-634-9205	Sequence 9205, Ap
41	43	44.3	4010	24	US-60-191-637-11384	Sequence 11384, A
42	43	44.3	4010	24	US-60-191-681-8915	Sequence 8915, Ap
43	43	43.3	123	1	PCT-US97-02218-350	Sequence 350, App
44	42	43.3	123	13	US-08-903-470-350	Sequence 350, App
45	42	43.3	181	1	PCT-US00-06112-1033	Sequence 1033, Ap

#### ALIGNMENTS

RESULT 1  
US-09-763-397A-3  
Sequence 3, Application US/09763397A  
GENERAL INFORMATION:  
APPLICANT: The Government of the United States of America, as represented by the  
APPLICANT: Secretary of the Department of Health and Human Services, Centers for  
APPLICANT: Control and Prevention  
APPLICANT: Lal, Altaf A.  
APPLICANT: Ping Shi, Ya  
APPLICANT: Hashmi, Seyed E.  
TITLE OF INVENTION: Recombinant Multivalent Malarial Vaccine Against Plasmodium Fa  
FILE REFERENCE: 6395-57049  
CURRENT APPLICATION NUMBER: US/09/763, 397A  
CURRENT FILING DATE: 2001-02-16  
PRIOR APPLICATION NUMBER: US 60/097,703  
PRIOR FILING DATE: 1998-08-21  
PRIOR APPLICATION NUMBER: PCT / US99/18869  
PRIOR FILING DATE: 1999-08-19  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 3  
LENGTH: 16  
TYPE: PRT  
ORGANISM: Plasmodium falciparum  
US-09-763-397A-3

Query Match 100.0%; Score 97; DB 21; Length 16;  
Best Local Similarity 100.0%; Pred. No. 1.7e-08;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 KPLDKFGNIYDYHEH 16



FILE REFERENCE: 9710-003-999  
CURRENT APPLICATION NUMBER: US/09/198,452A  
CURRENT FILING DATE: 1998-11-24  
NUMBER OF SEQ ID NOS: 6849  
SEQ ID NO 646  
LENGTH: 262  
TYPE: PRT  
ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-646

Query Match 49.5%; Score 48; DB 15; Length 262;  
Best Local Similarity 57.1%; Pred. No. 32;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 KPLDKFGNIYDYH 14  
DB 233 RPIDGFGNIRGIHY 246

RESULT 7  
US-09-438-185-608  
Sequence 608, Application US/09438185  
GENERAL INFORMATION:  
APPLICANT: Stephens, Richard  
APPLICANT: Mitchell, Wayne  
APPLICANT: Kalman, Sue  
APPLICANT: Davis, Ronald  
TITLE OF INVENTION: The Regents of the University of California  
FILE REFERENCE: 018941-000411US  
CURRENT APPLICATION NUMBER: US/09/438,185  
CURRENT FILING DATE: 1999-11-11  
PRIOR APPLICATION NUMBER: US 60/108,279  
PRIOR FILING DATE: 1998-11-12  
PRIOR APPLICATION NUMBER: US 60/128,606  
PRIOR FILING DATE: 1999-04-08  
NUMBER OF SEQ ID NOS: 1074  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 608  
LENGTH: 262  
TYPE: PRT  
ORGANISM: Chlamydia pneumoniae  
US-09-438-185-608

Query Match 49.5%; Score 48; DB 18; Length 262;  
Best Local Similarity 57.1%; Pred. No. 32;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 KPLDKFGNIYDYH 14  
DB 233 RPIDGFGNIRGIHY 246

RESULT 8  
US-09-328-352-7148  
Sequence 7148, Application US/09328352  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
FILE REFERENCE: GTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 7148  
LENGTH: 78  
TYPE: PRT  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-7148

Query Match 48.5%; Score 47; DB 17; Length 78;  
Best Local Similarity 56.2%; Pred. No. 12;  
Matches 9; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

QY 2 PLDKFGNIYDYH--YE 15  
DB 57 PLEKFAQLVDYHMAVE 72

RESULT 9  
US-09-328-352-5557  
Sequence 5557, Application US/09328352  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
FILE REFERENCE: GTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 5557  
LENGTH: 102  
TYPE: PRT  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-5557

Query Match 47.4%; Score 46; DB 17; Length 102;  
Best Local Similarity 58.3%; Pred. No. 23;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 PLDKFGNIYDYH 13  
DB 81 PLEKFAQLVDYH 92

RESULT 10  
US-09-328-352-5702  
Sequence 5702, Application US/09328352  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
FILE REFERENCE: GTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 5702  
LENGTH: 134  
TYPE: PRT  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-5702

Query Match 47.4%; Score 46; DB 17; Length 134;  
Best Local Similarity 58.3%; Pred. No. 32;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 PLDKFGNIYDYH 13  
DB 113 PLEKFAQLVDYH 124

RESULT 11  
US-09-328-352-7241  
Sequence 7241, Application US/09328352  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
FILE REFERENCE: GTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 7241  
; LENGTH: 141  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-7241

Query Match 47.4%; Score 46; DB 17; Length 141;  
Best Local Similarity 58.3%; Pred. No. 34;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 PLDKFGNIYDYH 13  
||:| : |||  
Db 120 PLEKFAQLVDYH 131

RESULT 12  
US-09-328-352-4183  
; Sequence 4183, Application US/09328352  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 4183  
; LENGTH: 146  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-4183

Query Match 47.4%; Score 46; DB 17; Length 146;  
Best Local Similarity 58.3%; Pred. No. 35;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 PLDKFGNIYDYH 13  
||:| : |||  
Db 125 PLEKFAQLVDYH 136

RESULT 13  
US-09-328-352-4906  
; Sequence 4906, Application US/09328352  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 4906  
; LENGTH: 146  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-4906

Query Match 47.4%; Score 46; DB 17; Length 146;  
Best Local Similarity 58.3%; Pred. No. 35;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 PLDKFGNIYDYH 13  
||:| : |||  
Db 125 PLEKFAQLVDYH 136

RESULT 14  
US-09-328-352-7059

; Sequence 7059, Application US/09328352  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 7059  
; LENGTH: 191  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-7059

Query Match 47.4%; Score 46; DB 17; Length 191;  
Best Local Similarity 58.3%; Pred. No. 48;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 PLDKFGNIYDYH 13  
||:| : |||  
Db 170 PLEKFAQLVDYH 181

RESULT 15  
US-09-328-352-7440  
; Sequence 7440, Application US/09328352  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 7440  
; LENGTH: 209  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-7440

Query Match 47.4%; Score 46; DB 17; Length 209;  
Best Local Similarity 58.3%; Pred. No. 53;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 PLDKFGNIYDYH 13  
||:| : |||  
Db 188 PLEKFAQLVDYH 199

Search completed: January 29, 2002, 10:55:59  
Job time: 2395 sec

---

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 10:58:07 ; Search time 120.95 Seconds  
(without alignments)  
4.841 Million cell updates/sec

Title: US-09-763-397A-3  
Perfect score: 97  
Sequence: 1 KPLDKRGNITYDYHEH 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 173191 seqs, 36597120 residues

Total number of hits satisfying chosen parameters: 173191

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA-New:\*  
1: /cgn2\_6/ptodata/2/paa/PCR\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep:\*  
7: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by the result to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	45.4	138	5	US-09-897-516-5215
2	44	45.4	459	5	US-09-815-242-13229
3	44	45.4	1404	5	US-09-708-427-21300
4	44	45.4	1420	5	US-09-708-427-11605
5	44	45.4	1517	5	US-09-708-427-11604
6	44	45.4	1519	5	US-09-708-427-11603
7	44	42.3	41	5	US-09-620-111B-8023
8	41	42.3	249	5	US-09-708-427-82968
9	41	42.3	359	5	US-09-708-427-65147
10	41	42.3	369	5	US-09-708-427-65146
11	41	42.3	395	5	US-09-708-427-65145
12	41	42.3	442	5	US-09-708-427-2795
13	41	42.3	497	5	US-09-708-427-2794
14	41	42.3	521	5	US-09-708-427-2793
15	40	41.2	212	5	US-09-708-427-68198
16	40	41.2	364	5	US-09-708-427-68197
17	39	40.2	121	6	US-10-015-127-11238
18	39	40.2	211	5	US-09-708-427-72557
19	39	40.2	262	5	US-09-708-427-67046
20	39	40.2	262	5	US-09-708-427-83736
21	39	40.2	279	5	US-09-708-427-67045
22	39	40.2	279	5	US-09-708-427-83735
23	39	40.2	303	5	US-09-708-427-67044
24	39	40.2	303	5	US-09-708-427-83734
25	39	40.2	393	5	US-09-708-427-17303
26	39	40.2	426	5	US-09-708-427-1466

27	39	40.2	447	5	US-09-708-427-17302	Sequence 17302, A
28	39	40.2	468	5	US-09-708-427-1465	Sequence 1465, Ap
29	39	40.2	481	5	US-09-708-427-17301	Sequence 17301, A
30	39	40.2	488	5	US-09-708-427-1464	Sequence 1464, Ap
31	39	40.2	1010	6	US-10-015-127-9620	Sequence 9820, Ap
32	39	40.2	1563	5	US-09-708-427-3644	Sequence 3644, Ap
33	39	40.2	1591	5	US-09-708-427-3643	Sequence 3643, Ap
34	39	40.2	1609	5	US-09-708-427-3642	Sequence 3642, Ap
35	38.5	39.7	172	5	US-09-985-153-91	Sequence 91, Appl
36	38.5	39.7	172	5	US-09-985-153-135	Sequence 135, Appl
37	38.5	39.7	267	5	US-09-985-153-70	Sequence 70, Appl
38	38	39.2	180	5	US-09-708-427-69129	Sequence 69129, A
39	38	39.2	187	5	US-09-708-427-80349	Sequence 80349, A
40	38	39.2	197	5	US-09-708-427-20520	Sequence 20520, A
41	38	39.2	207	5	US-09-708-427-20518	Sequence 20518, A
42	38	39.2	211	5	US-09-708-427-69128	Sequence 69128, A
43	38	39.2	218	5	US-09-708-427-80348	Sequence 80348, A
44	38	39.2	240	5	US-09-708-427-24797	Sequence 24797, A
45	38	39.2	255	5	US-09-708-427-24796	Sequence 24796, A

## ALIGNMENTS

```
RESULT 1
US-09-897-516-5215
; Sequence 5215 Application US/09897516
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Hinkle, Joseph E.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spiridonov, Sergei
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51847)B
; CURRENT APPLICATION NUMBER: US/09/897,516
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215, 161
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 5215
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Xenorhabdus sp.
US-09-897-516-5215

Query Match 45.4% Score 44: DB 5: Length 138:
Best Local Similarity 70.0% Pred. No. 4.9;
Matches 7: Conservative 0: Mismatches 3: Indels 0: Gaps 0:

QY 7 GNYDYHYEH 16
DB 28 GKAYDYHREH 37

RESULT 2
US-09-815-242-13229
; Sequence 13229 Application US/09815242
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes In
; PROKARYOTES
```

```
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 13229
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13229
```

```
Query Match          45.4%; Score 44; DB 5; Length 459;
Best Local Similarity 58.3%; Pred. No. 19;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
```

```
OY 5 KEGNIYDYHEH 16
    || || || |
Db 172 EFGYDYAHYH 183
```

```
RESULT 3
US-09-708-427-21300
; Sequence 21300, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21300
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..1404
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc.feature
; LOCATION: 1..1404
; OTHER INFORMATION: Ceres Seq. ID 1839339
US-09-708-427-21300
```

```
Query Match          45.4%; Score 44; DB 5; Length 1404;
Best Local Similarity 46.7%; Pred. No. 66;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
```

```
OY 1 KPLDKFGNIYDYHVE 15
    || | : || | : |
Db 102 KPSKKYKKLYDYFFE 116
```

```
RESULT 4
US-09-708-427-11605
; Sequence 11605, Application US/09708427
```

```
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11605
; LENGTH: 1420
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..1420
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc.feature
; LOCATION: 1..1420
; OTHER INFORMATION: Ceres Seq. ID 1823474
US-09-708-427-11605
```

```
Query Match          45.4%; Score 44; DB 5; Length 1420;
Best Local Similarity 46.7%; Pred. No. 67;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
```

```
OY 1 KPLDKFGNIYDYHVE 15
    || | : || | : |
Db 62 KPSKKYKKLYDYFFE 76
```

```
RESULT 5
US-09-708-427-11604
; Sequence 11604, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11604
; LENGTH: 1517
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..1517
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc.feature
; LOCATION: 1..1517
; OTHER INFORMATION: Ceres Seq. ID 1823473
US-09-708-427-11604
```

```
Query Match          45.4%; Score 44; DB 5; Length 1517;
Best Local Similarity 46.7%; Pred. No. 72;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
```

```
OY 1 KPLDKFGNIYDYHVE 15
    || | : || | : |
Db 159 KPSKKYKKLYDYFFE 173
```

```
RESULT 6
US-09-708-427-11603
; Sequence 11603, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
```

```
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11603
; LENGTH: 1519
; TYPE: PRF
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..1519
; OTHER INFORMATION: xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..1519
; OTHER INFORMATION: Ceres Seq. ID 1823472
US-09-708-427-11603
```

```
Query Match          45.4%; Score 44; DB 5; Length 1519;
Best Local Similarity 46.7%; Pred. No. 72;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 1 KPLDKFGNIYDYH 15
      |||::|||:|
DB 161 KPSKKYKKLYDFE 175
```

```
RESULT 7
US-09-620-111B-8023
; Sequence 8023, Application US/09620111B
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1070P
; CURRENT APPLICATION NUMBER: US/09/620,111B
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 9298
; SEQ ID NO 8023
; LENGTH: 41
; TYPE: PRF
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..41
; OTHER INFORMATION: xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..41
; OTHER INFORMATION: Ceres Seq. ID 1342121
US-09-620-111B-8023
```

```
Query Match          42.3%; Score 41; DB 5; Length 41;
Best Local Similarity 46.2%; Pred. No. 3.8;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 KPLDKFGNIYDYH 13
      |||::|||:|
DB 7 RPLDHPADLYDFH 19
```

```
RESULT 8
US-09-708-427-82968
; Sequence 82968, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
```

```
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 82968
; LENGTH: 249
; TYPE: PRF
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..249
; OTHER INFORMATION: xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..249
; OTHER INFORMATION: Ceres Seq. ID 1961854
US-09-708-427-82968
```

```
Query Match          42.3%; Score 41; DB 5; Length 249;
Best Local Similarity 46.2%; Pred. No. 29;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 KPLDKFGNIYDYH 13
      |||::|||:|
DB 7 RPLDHPADLYDFH 19
```

```
RESULT 9
US-09-708-427-65147
; Sequence 65147, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 65147
; LENGTH: 359
; TYPE: PRF
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..359
; OTHER INFORMATION: xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..359
; OTHER INFORMATION: Ceres Seq. ID 1928281
US-09-708-427-65147
```

```
Query Match          42.3%; Score 41; DB 5; Length 359;
Best Local Similarity 58.3%; Pred. No. 43;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 4 DKFGNIYDYH 15
      |||::|||:|
DB 77 DEYGYGYDYDE 88
```

```
RESULT 10
US-09-708-427-65146
; Sequence 65146, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 65146
```

```

?      LENGTH: 369
?      TYPE: PRT
?      ORGANISM: Zea mays subsp. mays
?      FEATURE:
?      NAME/KEY: misc_feature
?      LOCATION: 1..369
?      NAME/KEY: misc_feature
?      LOCATION: 1..369
?      OTHER INFORMATION: Ceres Seq. ID 1928280
US-09-706-427-65146

```

Query Match	42.3%	Score 41;	DB 5;	Length 369;
Best Local	58.3%	Pred No. 44;		
Matches	7;	Conservative	2;	Mismatches 3;
				Indels 0;
				Gaps 0

```

RESULT 11
US-09-708-427-65145
; Sequence 65145, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-1243p
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 65145
; LENGTH: 395
; TYPE: PRT
; ORGANSIM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..395
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..395
; OTHER INFORMATION: Ceres Seq. ID 1928279
; US-09-708-427-65145

```

Query Match	42.3%	Score 41	DB 5	Length 395
Best Local Similarity	58.3%	Pred. No. 48		
Matches	7	Conservative	2	Mismatches 3
				Indels 0
				Gaps 0

```

RESULT 12
US-09-708-427-2795
; Sequence 2795, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ. ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ. ID NO 2795
;
; LENGTH: 442
;
; TYPE: PRT
;
; ORGANISM: Arabidopsis thaliana

```

```

?      FEATURE:      misc_feature
?      NAME/KEY:     misc_feature
?      LOCATION:     1..442
?      OTHER INFORMATION: Xaa is any amino acid
?      NAME/KEY:     misc_feature
?      LOCATION:     1..442
?      OTHER INFORMATION: Ceres Seq. ID 1809716
?      OS-09-708-427-2195

```

Query Match	42.3%	Score 41	DB 5	Length 442
Best Local	Similarity 53.8%	Pred. NO. 54		
Matches	7	Conservative	2	Mismatches 4
				Indels 0
				Gaps 0

```

RESULT 13
US-09-708-427-2794
; Sequence 2794, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2794
; LENGTH: 497
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..497
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc.feature
; LOCATION: 1..497
; OTHER INFORMATION: Ceres Seq. ID 1809715
; US-09-708-427-2794

```

Query Match	42.3%;	Score 41;	DB 5;	Length 497;
Best Local Similarity	53.8%;	Pred. No. 62;		
Matches	7;	Conservative	2;	Mismatches 4; Indels 0; Gaps 0

```

RESULT 14
US-09-708-427-2793
; Sequence 2793, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2793
; LENGTH: 521
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..521

```

```

; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..521
; OTHER INFORMATION: Ceres Seq. ID 1809714
US-09-708-427-2793

```

```

Query Match          42.3%; Score 41; DB 5; Length 521;
Best Local Similarity 53.8%; Pred. No. 65;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```

```

OY 1 KPLDKFGNIYDYH 13
    |||| : |||
Db 265 KPLDHPADYFDYH 277

```

```

RESULT 15
US-09-708-427-68198
; Sequence 68198, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ. ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 68198
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..212
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..212
; OTHER INFORMATION: Ceres Seq. ID 1934091
US-09-708-427-68198

```

```

Query Match          41.2%; Score 40; DB 5; Length 212;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 4 DKFGNIY 10
    |||||
Db 44 DKFGNIY 50

```

Search completed: January 29, 2002, 10:58:07  
 Job time: 2503 sec

**THIS PAGE BLANK (USPTO)**

	FT	Region	/label= artifact 104
	FT		/note= "see comments"
	FT	Region	105..209
	PT		/label= AAs 19-123 of CS protein
	FT		/note= "Region I contg. flanking region less signal sequence"
	PT	Region	210
	FT		/label= artifact
	PT		/note= "see comments"
	FT	Region	211..335
	PT		/label= AAs 288-412 of CS protein
	FT		/note= "Region II flanking region"
PB	XN		
PD	XK	EP432965-A.	
		19-JUN-1991.	
PF	KX	06-DEC-1990;	90EP-0313257.
PK	XK	08-DEC-1989;	89US-0447746.
PA	(SMIK ) SWITZERLINE BEECHAM.		
PA	(USSA ) US SEC OF THE ARMY.		
PA	(BIOM-) BIOMEDICAL RES INST.		
PI	Gross MS; Gordon DM; Hollingdale MR;		
DR	WFI; 1991-179771/25.		
XX			
PS			
XX	Example 4; Page 11; 18pp; English.		
CC	The polypeptide comprising immunogenic determinants from P falciparum		
CC	- for vaccine against malaria infection in humans.		
CC			
CC	Science 225 : 593 (1984)), and the influenza virus non-structural		
CC	protein 1 (NSI) [Baes et al., Nucleic Acids Research, 8 : 5845		
CC	(1980)]. The DNA encoding the 1st 61 AAs of the N-terminal of NSI		
CC	(NSI_81) is linked to a synthetic sequence encoding four repeat		
CC	units from the immunodominant region, which in turn is linked via		
CC	a synthetic sequence to DNA encoding Region I contg. flanking		
CC	region less the 18 AA signal region. This is linked to DNA		
CC	encoding Region II-contg. flanking region. The pro. residue sep-		
CC	erating the Asp (at the C-terminal of the linker) from the Region		
CC	I-contg. CS flanking region is an artifact of a filled-in BamHI		
CC	site; the Gly separating the Region I and II-contg. CS flanking		
CC	regions can be used as an artifact of a synthetic FokI/PvuII linker. The		
CC	peptide can be utilized in a vaccine for protection against malaria.		
CC	See also AAR12306-R12311 and AAR13175-R13179.		
SQ			
Sequence	335 AA:		
Oy	Query Match	100.0%;	Score 127;
Dz	Best Local Similarity	100.0%;	Pred. No. 1.7e+10;
Matches	23; Conservative	0;	Mismatches 0;
		Indels 0;	Gaps 0;
	1 KPKDELDYNDIEKKICKMEKS 23		
	DB 291 kpkdeldyndiekkickmekcs 313		
RESULT	B		
ID	AAR13179 standard; Protein; 335 AA.		
AC	AAR13179;		
TX	29-AUG-1991 (first entry)		
DE	NSL_81(NVDP)ARLFauth.		

Key	Location/Qualifiers
Region	1..81
Region	/label="N-terminal of NS1"
Region	/note="Influenza virus nonstructural protein 1"
Region	82..97
Region	/label="Immunodominant repeat region"
Region	/note="Four variant tetrapeptide repeat units"
Region	98..103
Region	/label="synthetic linker"
Region	104
Region	/label="artifact"
Region	/note="see comments"
Region	105..209
Region	/label="AAs 19-123 of CS protein"
Region	/note="1 contig, flanking region less signal sequence"
Region	210
Region	/label="artifact"
Region	/note="see comments"
Region	211..335
Region	/label="AAs 288-412 of CS protein"
Region	/note="Region II flanking region"
Region	EP432965-A.
Region	19-JUN-1991.
Region	06-DEC-1990. 90EP-033257.
Region	08-DEC-1989. 8905-0447746.
Region	(SMIK ) SMITHKLINE BEECHAM.
Region	(USA ) US SEC OF THE ARMY.
Region	(BIOM-) BIOMEDICAL RES INST.
Region	Gross MS, Gordon DM, Hollingdale MR;
Region	WPI, 1991-179771/25.
Region	Polypeptide comprising immunogenic determinants from P falciparum
Region	- for vaccine against malaria infection in humans.
Region	Example 5; Page 11; 18pp; English.
Region	The polypeptide is prep'd. by genetic engineering of genes encoding
Region	the P. falciparum circumsporozoite (CS) protein [Bame et al.,
Region	Science 225 : 593 (1984)], and the Influenza virus non-structural
Region	protein 1 (NS1), [Baez et al., Nucleic Acids Research, 8 : 5845
Region	(1980)]. The DNA encoding the 1st 81 AAs of the N-terminal of NS1
Region	(NS1.81) is linked to a synthetic sequence encoding four repeat
Region	units (the variant form) from the immunodominant region, which in
Region	turn is linked via a synthetic sequence to DNA encoding Region I
Region	contg. flanking region less the 18 AA signal region. This is
Region	linked to DNA encoding Region II-contg. flanking region. The Pro
Region	residue separating the Asp (at the C-terminal of the linker) from
Region	the Region I-contg. CS flanking region is an artifact of a filled
Region	in BamHI site: the Gly separating the Region I and II-contg. CS
Region	flanking regions is an artifact of a synthetic PstI/ThiI I
Region	linker. The peptide can be used in a vaccine for protection
Region	against malaria.
Region	See also AAR13506-R12311 and AAR13175-R13178.
Region	Sequence 335 AA.

```
Query Match      100.0%; Score 127; DB 12; Length 335;
```

Best Local Similarity 100.0%; Pred. No. 1,7e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPKDELVDYNDIEKKICKMEKS 23  
DB 291 KPKDELVDYNDIEKKICKMEKS 313

## RESULT 9

AA70278  
ID AAY70278 standard; protein; 350 AA.

AC AAY70278;

DT 06-JUN-2000 (first entry)

DE Recombinant vaccine CDC/NIIMALVAC-1.

KM Recombinant protein: CDC/NIIMALVAC-1; multivalent; malarial; vaccine;  
KM T-cell epitope; tetanus toxoid; antigenic epitope; treatment; vaccine;  
KM Circumsporozoite protein; CSP; sporozoite surface protein-2; SSP-2;  
KM Liver stage antigen-1; LSA-1; merozoite surface protein-1; MSP-1;  
KM Apical membrane antigen-1; AMA-1; erythrocyte binding antigen-175;  
KM EBA-175; rhoptry associated protein-1; RAP-1; gamete specific antigen;  
KM Pf27; antiparasitic; prevention; anti-CDC/NIIMALVAC-1 antibody;  
KM honey bee.

OS Chimeric - Apis sp.

OS Chimeric - Clostridium tetani.

OS Chimeric - Plasmodium falciparum.

PH Key location/Qualifiers

FT Peptide

FT 1..22 /label= Melittin-signal-peptide

FT /note= "Derived from Honey Bee"

FT Protein

FT 23..350 /label= Mature\_CDC/NIIMALVAC-1

FT /note= "Recombinant multivalent malarial vaccine"

PN WO200011179-A1.

PD 02-MAR-2000.

PF 19-AUG-1999; 99WO-US18669.

PR 21-AUG-1998; 98US-0097703.

PA (NATM-) NAT INST IMMUNOLOGY.

PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.

PI Tal AA, Shi YP, Hasnain SE;

DR WPI: 2000-237654/20.

DR N-PSDB; AAZ51336.

PT Novel recombinant protein as vaccine for treating malarial infection

PT comprises antigenic peptides obtained from different stages of

PT plasmodium falciparum life cycle. -

PS Claim 3; Page 43-44; 52pp; English.

CC The present sequence is that of recombinant protein CDC/NIIMALVAC-1,  
CC which is a multivalent, multistage malarial vaccine. The recombinant  
CC protein comprises, melittin signal peptide, (His)6 tag, T-cell epitope  
CC from tetanus toxoid and 21 antigenic epitopes from circumsporozoite  
CC protein (CSP), sporozoite surface protein-2 (SSP-2), liver stage  
CC antigen-1 (LSA-1), merozoite surface protein-1 (MSP-1), MSP-2, apical  
CC membrane antigen-1 (AMA-1), erythrocyte binding antigen-175 (EBA-175),  
CC rhoptry associated protein-1 (RAP-1) and gamete specific antigen, Pf27.  
CC These epitopes were obtained at different stages of the life cycle of  
CC plasmodium falciparum. CDC/NIIMALVAC-1 vaccine has antiparasitic  
CC activity and can be used for treatment and prevention of malarial  
CC infections. Anti-CDC/NIIMALVAC-1 antibodies can be used for detecting

CC P. falciparum in biological samples.

SQ Sequence 350 AA;

Query Match 100.0%; Score 127; DB 21; Length 350;

Best Local Similarity 100.0%; Pred. No. 1.8e-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPKDELVDYNDIEKKICKMEKS 23

DB 51 KPKDELVDYNDIEKKICKMEKS 73

## RESULT 10

AA83144  
ID AAP83144 standard; protein; 411 AA.

AC AAP83144;

DT 20-NOV-1990 (first entry)

DE Sequence encoded by the circumsporozoite (CS) gene from

DE Plasmodium falciparum.

KM Vaccine; antigen; immunogen; probe; hybridisation;

KM immunosay; diagnosis.

OS Plasmodium falciparum.

PH Key location/Qualifiers

FT Region 106..120

FT /note= "Region 1"

FT Region 147..206

FT /note= "Repeat region, repeat unit = NAMP"

FT Region 123..146

FT /note= "Repeat region, repeat unit = NAMPNDP"

FT Region 211..286

FT /note= "Repeat region, repeat unit = NAMP"

PN EP278940-A.

PD 17-AUG-1988.

PF 25-JAN-1988; 88EP-0870008.

PR 30-JAN-1987; 87US-0009325.

PA (SMIK ) SMITH KLINE-RIT.

PI Cabazon T, De Wilde M, Harford N;

DR WPI: 1988-229751/33.

DR N-PSDB; AAN81108.

PT DNA encoding hepatitis B virus antigens and hybrids contg. them -

PT used for expression in yeast to obtain vaccines and bivalent

PT vaccines

PS Example; Fig 3Aa-3Af; 101pp; English.

CC Sequence of the CS gene (AAN81108) is from lambda-mpf1. A recombinant

CC DNA molecule is claimed, comprising functional DNA coding sequence fused,

CC in phase, to a portion of the Pre-S2 region of a hepatitis B virus (HBV)

CC Pre-S2 protein coding sequence. The functional DNA coding sequence

CC comprises the Pre-S2 coding sequence, the Pre-S1 coding sequence or entire

CC Pre-S1-Pre-S2 coding sequence, the CS protein coding sequence or

CC an HIV envelope gene sequence, e.g. HIV C7 protein coding region, HIV

CC Peptide 121 coding region, or HIV Dreesman peptide coding region.



RA Olafsson P., Matile H., Certa U.;  
RT "Plasmodium falciparum: the repetitive MSA-1 surface protein of the  
RT RO-71 isolate is recognized by mouse antibody against the  
RT nonrepetitive repeat block of RO-33.";  
RL Exp. Parasitol. 74:381-389(1992).  
DR EMBL: X61930; CAA43932.1;  
DR InterPro: IPR000087; Collagen.  
FT NON\_TER 1087 1087  
SQ SEQUENCE 1087 AA; 123911 MW; 5518852133C01B33 CRC64;

Query Match 57.4%; Score 42.5; DB 5; Length 1087;  
Best Local Similarity 60.0%; Pred. No. 53;  
Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

OY 1 QYIKANSKFI-GITE 14  
DB 1009 QFVKSNSKVITGLTE 1023

RESULT 13  
O9T2T5 PRELIMINARY; PRT; 1694 AA.  
AC O9T2T5;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR.  
GN MSP1.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HN1;  
RA Jiang G., Liu R.Z., Daubenberger C.A., Pluschke G.;  
RL Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF062348; AAC72884.1;  
DR InterPro: IPR000561; EGF-like.  
DR Pfam: PF00008; EGF; 1.  
KW Merozoite.  
SQ SEQUENCE 1694 AA; 192794 MW; 84CF0E709F5673B CRC64;

Query Match 57.4%; Score 42.5; DB 5; Length 1694;  
Best Local Similarity 60.0%; Pred. No. 84;  
Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

OY 1 QYIKANSKFI-GITE 14  
DB 995 QFVKSNSKVITGLTE 1009

RESULT 14  
O9NHX1 PRELIMINARY; PRT; 1694 AA.  
AC O9NHX1;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE MAJOR MEROZOITE SURFACE ANTIGEN.  
GN GPI95.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FC1/HN;  
RA Shan Z.X., Yu X.B., Li X.R., Ma C.L., Fang J.M.;  
RT "Molecular cloning and sequence analysis of major merozoite surface  
RT antigen(gp195)gene of Plasmodium falciparum isolate FCC1/HN.";  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF218248; AAF27526.1;

DR InterPro: IPR000561; EGF-like.  
DR Pfam: PF00008; EGF; 1.  
KW Merozoite.  
SQ SEQUENCE 1694 AA; 192766 MW; B51634A49E0F6728 CRC64;

Query Match 57.4%; Score 42.5; DB 5; Length 1694;  
Best Local Similarity 60.0%; Pred. No. 84;  
Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

OY 1 QYIKANSKFI-GITE 14  
DB 995 QFVKSNSKVITGLTE 1009

RESULT 15  
O9T2T4 PRELIMINARY; PRT; 1704 AA.  
AC O9T2T4;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR.  
GN MSP1.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HN2;  
RA Jiang G., Liu R.Z., Daubenberger C.A., Pluschke G.;  
RL Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF062349; AAC72885.1;  
DR InterPro: IPR000561; EGF-like.  
DR Pfam: PF00008; EGF; 1.  
KW Merozoite.  
SQ SEQUENCE 1704 AA; 193762 MW; 38526DD0DA56FD1D CRC64;

Query Match 57.4%; Score 42.5; DB 5; Length 1704;  
Best Local Similarity 60.0%; Pred. No. 85;  
Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

OY 1 QYIKANSKFI-GITE 14  
DB 1005 QFVKSNSKVITGLTE 1019

Search completed: January 29, 2002, 11:15:53  
Job time: 213 sec



ID 09XRJ9 PRELIMINARY: PRT: 71 AA.  
AC 09XRJ9  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE MHC CLASS II BETA 1 (FRAGMENT).  
GN SANA.  
OS Salvelinus namaycush (lake trout).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salvelinus.  
OX NCBI\_TaxID=8040.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Dorschner M.O., Duris T., Phillips R.B.;  
RT "Diversity of a lake trout Mhc class II gene."  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- MISCELLANEOUS: FOUND N-TERMINAL TO IMMUNOGLOBULIN AND MAJOR  
HISTOCOMPATIBILITY COMPLEX DOMAIN (BY SIMILARITY).  
CC -1- SIMILARITY: TO CLASS II HISTOCOMPATIBILITY ANTIGEN, BETA CHAIN.  
DE BETA-1 DOMAIN INTERPRO FAMILY.  
DR EMBL: AF130026; AAD20889.1; -  
DR InterPro: IPR000353; MHC\_II\_beta.  
DR Pfam: PF00969; MHC\_II\_beta.1.  
DR ProDom: PD000328; MHC\_II\_beta.1.  
DR Glycoprotein: MHC II; Transmembrane.  
FT NON\_TER 1  
KW SEQUENCE 71 AA; 8002 MW; E3095286582A9F2D CRC64;  
SO  
Query Match 60.1%; Score 44.5; DB 7; Length 71;  
Best Local Similarity 55.6%; Pred. No. 1.4;  
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;  
OY 1 OYIKANS---KFIGITEL 15  
Db 14 EYIRNSTVGKFGVGYTEL 31  
RESULT 3  
ID 031590 PRELIMINARY: PRT: 244 AA.  
AC 031590  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE MHC CLASS II.  
OS Salmo salar (Atlantic salmon).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.  
OX NCBI\_TaxID=8030.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA TISSUE-LEUKOCYTES;  
RC Hotdvalik I., Grimholt U., Fosse V.M., Lie Y., Endresen C.;  
RL Immunogenetics 0:0-0(0).  
CC -1- MISCELLANEOUS: FOUND N-TERMINAL TO IMMUNOGLOBULIN AND MAJOR  
HISTOCOMPATIBILITY COMPLEX DOMAIN (BY SIMILARITY).  
CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX  
DOMAIN.  
CC -1- SIMILARITY: TO CLASS II HISTOCOMPATIBILITY ANTIGEN, BETA CHAIN.  
DE BETA-1 DOMAIN INTERPRO FAMILY.  
DR EMBL: X70166; CAA49725.1; -  
DR InterPro: IPR003597; Ig\_cl.  
DR InterPro: IPR003006; Ig\_MHC.  
DR InterPro: IPR000353; MHC\_II\_beta.  
DR Pfam: PF00047; Ig.1.  
DR Pfam: PF00969; MHC\_II\_beta.1.  
DR ProDom: PD000328; MHC\_II\_beta.1.  
DR SMART: SM00407; Igcl.1.  
KW Glycoprotein: MHC; MHC II; Transmembrane.  
FT NON\_TER 1

SO SEQUENCE 244 AA; 27449 MW; 496CB9EA9D73765C CRC64;  
Query Match 60.1%; Score 44.5; DB 7; Length 244;  
Best Local Similarity 55.6%; Pred. No. 5;  
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;  
OY 1 OYIKANS---KFIGITEL 15  
Db 51 EYIRNSTVGKFGVGYTEL 68  
RESULT 4  
ID 09XG37 PRELIMINARY: PRT: 546 AA.  
AC 09XG37  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE HYPOTHETICAL 66.2 KDA PROTEIN.  
OS Guillardia theta (Cryptomonas phi).  
OC Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.  
OX NCBI\_TaxID=55529.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Zauner S., Fraunholz M., Wastl J., Penny S.L., Cavaller-Smith T.,  
RA Maier U., Douglas S.;  
RT "Aberrant telomeres, overlapping genes and chloroplast protein-  
coding functions in an unusually compact eukaryotic genome - the  
cryptomonad nucleomorph."  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ010592; CAB40403.1; -  
KW Hypothetical protein.  
SO SEQUENCE 546 AA; 66218 MW; 7303950F632BE6F2 CRC64;  
Query Match 59.5%; Score 44; DB 10; Length 546;  
Best Local Similarity 50.0%; Pred. No. 14;  
Matches 7; Conservative 6; Mismatches 1; Indels 0; Gaps 0;  
OY 2 YIKANSKFIGITEL 15  
Db 445 FIKSNSRFRRLTEI 458  
RESULT 5  
ID 031578 PRELIMINARY: PRT: 67 AA.  
AC 031578  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE (DB03) MHC CLASS II BETA 1 (FRAGMENT).  
OS Salmo salar (Atlantic salmon).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.  
OX NCBI\_TaxID=8030.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Grimholt U., Olaker I., Vries Lindstrom C., Lie O.;  
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.  
CC -1- MISCELLANEOUS: FOUND N-TERMINAL TO IMMUNOGLOBULIN AND MAJOR  
HISTOCOMPATIBILITY COMPLEX DOMAIN (BY SIMILARITY).  
CC -1- SIMILARITY: TO CLASS II HISTOCOMPATIBILITY ANTIGEN, BETA CHAIN.  
DE BETA-1 DOMAIN INTERPRO FAMILY.  
DR EMBL: L24929; AAA49590.1; -  
DR InterPro: IPR000353; MHC\_II\_beta.  
DR Pfam: PF00969; MHC\_II\_beta.1.  
DR ProDom: PD000328; MHC\_II\_beta.1.  
KW Glycoprotein: MHC; MHC II; Transmembrane.  
FT NON\_TER 1  
FT NON\_TER 67

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 11:15:52 ; Search time 62.77 Seconds  
(without alignments)  
34.954 Million cell updates/sec

Title: US-09-763-397a-24  
Perfect score: 74  
Sequence: 1 OYIKANSKFTGTEL 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP-archaea:\*  
2: SP-bacteria:\*  
3: SP-fungi:\*  
4: SP-human:\*  
5: SP-invertebrate:\*  
6: SP-mammal:\*  
7: SP-mhc:\*  
8: SP-phage:\*  
9: SP-plant:\*  
10: SP-rodent:\*  
11: SP-virus:\*  
12: SP-vertebrate:\*  
13: SP-unclassified:\*  
14: SP-unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44.5	60.1	60	7	031585
2	44.5	60.1	71	7	09XRJ9
3	44.5	60.1	244	7	031590
4	44	59.5	546	10	09XG37
5	43.5	58.8	67	7	031578
6	43	58.1	180	2	09CF66
7	43	58.1	250	9	09MCL7
8	43	58.1	252	9	09XJF8
9	43	58.1	302	11	09CRV4
10	43	58.1	309	11	09CYD2
11	43	58.1	899	12	09YTK4
12	42.5	57.4	1087	5	025961
13	42.5	57.4	1694	5	09TRT5
14	42.5	57.4	1694	5	09NHX1
15	42.5	57.4	1704	5	09TRT4
16	42.5	57.4	1720	5	025922
17	42	56.8	1333	5	024262
18	41.5	56.1	84	13	09DEK4
19	41.5	56.1	149	7	031495

20	41.5	56.1	216	7	09GJH0	09gjh0 salmo trutt
21	41.5	56.1	216	7	09GJG9	09gjb9 salmo trutt
22	41	55.4	247	11	09B3B9	09db07 mus musculu
23	41	55.4	384	4	09BD07	09hd07 homo sapien
24	41	55.4	532	5	096671	096671 drosophila
25	41	55.4	540	5	09VU53	09vut3 drosophila
26	40.5	54.7	67	7	031577	031577 salmo salar
27	40.5	54.7	67	7	031581	031581 salmo salar
28	40.5	54.7	67	7	031582	031582 salmo salar
29	40.5	54.7	71	7	09XRH6	09xrth6 salmo salar
30	40.5	54.7	71	7	09XRH1	09xrth1 salmo salar
31	40.5	54.7	84	13	09DEJ6	09dej6 coregonus s
32	40.5	54.7	216	7	09GJH2	09gjh2 salmo trutt
33	40.5	54.7	217	7	09GJH5	09gjh5 salmo trutt
34	40.5	54.7	245	7	031591	031591 salmo salar
35	40	54.1	134	11	09DA16	09da16 mus musculu
36	40	54.1	601	2	09CEV7	09cev7 lactococcus
37	39.5	53.4	71	7	09XRH4	09xrth4 salmo salar
38	39.5	53.4	131	7	046866	046866 ictalurus p
39	39.5	53.4	131	7	046869	046869 ictalurus p
40	39.5	53.4	217	7	09GJH7	09gjh7 salmo trutt
41	39.5	53.4	217	7	09GJH6	09gjh6 salmo trutt
42	39	52.7	131	12	09YPT2	09ypt2 ipomoea yel
43	39	52.7	167	4	09BYL3	09byl3 homo sapien
44	39	52.7	304	4	09UBP9	09ubp9 homo sapien
45	39	52.7	326	2	09ZJ72	09zj72 helicobacte

ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	60 AA
1	031585			
AC	031585			
DT	01-NOV-1996 (TRENBLREL. 01, Created)			
DT	01-NOV-1996 (TRENBLREL. 01, Last sequence update)			
DT	01-JUN-2001 (TRENBLREL. 17, Last annotation update)			
DE	(DB02) MHC CLASS II BETA 1 (FRAGMENT).			
OS	Salmo salar (Atlantic salmon).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei;			
OC	Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.			
OX	NCBI_TaxID=8030;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Grimholt U., Olsaker I., Vries Lindstrom C., Lie O.,			
RL	Submitted (Oct-1993) to the EMBL/GenBank/DBD databases.			
CC	-1- MISCELLANEOUS: FOUND N-TERMINAL TO IMMUNOGLOBULIN AND MAJOR			
CC	HISTOCOMPATIBILITY COMPLEX DOMAIN (BY SIMILARITY).			
CC	-1- SIMILARITY: TO CLASS II HISTOCOMPATIBILITY ANTIGEN, BETA CHAIN,			
CC	BETA-1 DOMAIN INTERPRO FAMILY.			
DR	EMBL: L24953; AAA49597.1; -			
DR	InterPro: IP000353; MHC_II_beta.			
DR	Pfam: PF00969; MHC_II_beta.1.			
DR	ProDom: PD000328; MHC_II_beta.1.			
KW	Glycoprotein; MHC; MHC II; Transmembrane.			
FT	NON_TER 1			
FT	NON_TER 60			
FT	NON_TER 60			
SQ	SEQUENCE 60 AA; 6776 MW; 5D4F3449060940E2 CRC64;			

Query Match	60.1%	Score 44.5;	DB 7;	Length 60;
Best Local Similarity	55.6%	Pred. No. 1.2;		
Matches 10;	Conservative	3;	Mismatches	2;
			Indels	3;
			Gaps	1;
OY	1 OYIKANS--KFTGTEL 15			
	:11:11:11:11:11			
DB	16 EYIRPNSWTKKFTGTEL 33			
RESULT	2			
09XRJ9				

XX Novel recombinant protein as vaccine for treating malarial infection  
PT comprises antigenic peptides obtained from different stages of  
plasmidium falciparum life cycle  
XX  
XX Claim 2; Page 16; 52pp; English.  
XX  
XX The present sequence is the antigenic epitope P593, derived from  
CC circumsporozoite protein (CSP) of the sporozoite stage of Plasmodium  
CC falciparum. It is used in the construction of recombinant protein  
CC COC/NIMALVAC-1, which is a multivalent, multistage malarial vaccine.  
CC The recombinant protein comprises, melittin signal peptide, (His)6 tag,  
CC T-cell epitope from tetanus toxoid and 21 antigenic epitopes from  
CC circumsporozoite protein (CSP), sporozoite surface protein-2 (SSP-2),  
CC liver stage antigen-1 (LSA-1), merozoite surface protein-1 (MSP-1),  
CC MSP-2, apical membrane antigen-1 (AMA-1), erythrocyte binding antigen-175  
CC (EBA-175), rhoptry associated protein-1 (RAP-1) and gamete specific  
CC antigen, Pf27. These epitopes were obtained at different stages of the  
CC life cycle of P. falciparum. CDC/NIMALVAC-1 vaccine has antiparasitic  
CC activity and can be used for treatment and prevention of malarial  
CC infections. Anti-CDC/NIMALVAC-1 antibodies can be used for detecting  
XX P. falciparum in biological samples.  
XX  
XX Sequence 23 AA;

Query Match 100.0%; Score 127; DB 21; Length 23;  
Best Local Similarity 100.0%; Pred. No. 8.4e-12;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPKDELDYENDIEKKICKMEKCS 23  
|||||  
DB 1 kpkdelyendiekkickmekcs 23

## RESULT 2

ID AAR66442 standard; peptide; 66 AA.  
XX  
XX AAR66442;  
XX  
XX 03-AUG-1995 (first entry)  
XX  
XX Plasmodium falciparum circumsporozoite antigen.  
XX  
XX T cell helper site; cytotoxic T cell response; neutralising antibody;  
KW P.falciparum CS antigen; circumsporozoite; malaria; vaccine.  
XX  
XX Synthetic.  
XX  
XX Key Location/Qualifiers  
XX Region 22..41  
FT /label= repeat\_region  
FT /note= "(NANP)5"  
XX  
XX WO9426785-A.  
XX  
XX 24-NOV-1994.  
XX  
XX 13-MAY-1994; 94WO-US05142.  
XX  
XX 14-MAY-1993; 93US-0060988.  
XX  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX (USSH ) US SEC DEPT HEALTH.  
XX  
XX Ahlers JD, Berzofsky JA, Nara P, Pendleton CD, Shirai M;  
XX WPI; 1995-006707/01.  
XX  
XX Polypeptide inducing helper T cell, cytotoxic T cell and  
PT antibodies responses - to target antigen in hosts of different  
PT MHC haplotypes, esp. for therapeutic or prophylactic vaccines

PT against HIV.  
XX  
XX Example 4; Page 54; 120pp; English.  
XX  
XX The peptide AAR66442 was synthesised to make a peptide vaccine which  
CC is recognised by a broad range of MHC types that will elicit a T  
CC helper cell response, a CTL response and a high titre neutralising  
CC antibody response.  
XX  
XX Sequence 66 AA;

Query Match 100.0%; Score 127; DB 16; Length 66;  
Best Local Similarity 100.0%; Pred. No. 2.7e-11;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPKDELDYENDIEKKICKMEKCS 23  
|||||  
DB 44 kpkdelyendiekkickmekcs 66

## RESULT 3

ID AAR13175 standard; Protein; 309 AA.  
XX  
XX AAR13175;  
XX  
XX 29-AUG-1991 (first entry)  
XX  
XX NS1\_81-RLfdelta9.  
XX  
XX Immunogenic determinant; circumsporozoite; CS; vaccine; malaria;  
KW hybrid; influenza virus; non-structural protein 1; fusion.  
XX  
XX Plasmodium falciparum.  
OS Influenza virus (A/PR/8/34/).  
XX  
XX Key Location/Qualifiers  
XX Region 1..81  
FT /label= N-terminal of NS1  
FT /note= "Influenza virus nonstructural protein 1"  
FT Peptide 82..87  
FT /label= synthetic linker  
FT Region 88  
FT /label= artifact  
FT /note= "see comments"  
FT Region 89..193  
FT /label= AAs 19-123 of CS protein  
FT /note= "Region 1 contg. flanking region less  
FT signal sequence"  
FT Region 194..309  
FT /label= AAs 297-412 of CS protein  
FT /note= "Region II flanking region minus 9 N-term-  
FT inal AAs"  
XX  
XX EP432965-A.  
XX  
XX 19-JUN-1991.  
XX  
XX 06-DEC-1990; 90EP-0313257.  
XX  
XX 08-DEC-1989; 89US-0447746.  
XX  
XX (SMIK ) SMITHKLINE BEECHAM.  
XX (USSA ) US SEC OF THE ARMY.  
XX (BIOM-) BIOMEDICAL RES INST.  
XX  
XX Gross MS, Gordon DM, Hollingdale MR;  
XX WPI; 1991-179771/25.  
XX  
XX Polypeptide comprising immunogenic determinants from P falciparum  
PT - for vaccine against malaria infection in humans.

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 10:21:42 ; Search time 310.82 Seconds  
(without alignments)  
5.481 Million cell updates/sec

Title: US-09-763-397A-6

Perfect score: 127

Sequence: 1 RPKDELVDYENDIEKKICKMEKCS 23

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_1101.\*  
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	127	100.0	23	21 AAY70282	Plasmodium falciparum
2	127	100.0	66	16 AAR66442	Plasmodium falciparum
3	127	100.0	309	12 AAR13175	NS1_81-RLFAuth.9
4	127	100.0	319	11 AAR07945	NS181RLFAuth plasm
5	127	100.0	319	12 AAR13176	NS1_81-RLFAuth. P
6	127	100.0	327	12 AAR13177	NS1_81-RLFAuth. P
7	127	100.0	335	12 AAR13178	NS1_81(NANP)4RLFAu
8	127	100.0	335	12 AAR13179	NS1_81(NVDP)4RLFAu
9	127	100.0	350	21 AAY70278	Recombinant vaccin
10	127	100.0	411	9 AAR83144	Sequence encoded b
11	127	100.0	412	7 AAR60416	CS protein of mala

12	127	100.0	424	14 AAR37796	RTS protein. Synt
13	123	96.9	412	9 AAR80835	Sequence encoded b
14	121	95.3	424	14 AAR37797	RTS* protein. Syn
15	120	94.5	180	11 AAR07290	Circumsporozoite a
16	96	75.6	40	16 AAR87213	P.falciparum deriv
17	75	59.1	20	16 AAR71651	CS Region II+ mine
18	71	55.9	21	21 AAY70283	Plasmodium falcipa
19	62	48.8	14	16 AAR71655	Circumsporozoite a
20	62	48.8	19	7 AAR61002	Sequence corresp.
21	60.5	47.6	402	8 AAR70709	Plasmodium cynomol
22	58.5	46.1	19	16 AAR71654	CS Region II+ mine
23	58.5	46.1	168	11 AAR07291	Circumsporozoite a
24	58.5	46.1	272	10 AAR93560	Plasmodium berghei
25	56.5	44.5	478	21 AAR07288	Human II-2/ P. yoe
26	53	41.7	19	21 AAY99033	H1A class II bindi
27	53	41.7	20	22 AAY04118	P.falciparum CS pr
28	53	41.7	21	10 AAR91504	Sequence of modifi
29	53	41.7	21	16 AAR82586	Plasmodium falcipa
30	53	41.7	21	16 AAR78920	Malaria circumspor
31	53	41.7	21	16 AAR75955	P. falciparum CS p
32	53	41.7	21	16 AAR70912	Malaria circumspor
33	53	41.7	21	17 AAR05612	Circumsporozoite h
34	53	41.7	21	18 AAR35440	T-cell stimulatory
35	53	41.7	21	20 AAY23252	Peptide derived fr
36	53	41.7	21	21 AAY80071	Pathogen derived T
37	53	41.7	21	21 AAY54553	T helper cell (Th)
38	53	41.7	21	21 AAY58777	Unidentified pepti
39	53	41.7	21	22 AAB99706	Plasmodium falcipa
40	53	41.7	21	22 AAG62428	Plasmodium falcipa
41	53	41.7	21	22 AAG84517	Plasmodium falcipa
42	53	41.7	21	22 AAG88269	Plasmodium falcipa
43	53	41.7	21	22 AAG89366	Plasmodium falcipa
44	53	41.7	21	22 AAB84447	Sequence of T help
45	53	41.7	21	22 AAB98457	Plasmodium falcipa

## ALIGNMENTS

RESULT 1  
AAY70282  
ID AAY70282 standard; peptide; 23 AA.  
XX  
AC AAY70282;  
XX  
DT 06-JUN-2000 (first entry)  
XX  
DE Plasmodium falciparum CSP antigenic epitope, P593.  
XX  
KW Recombinant protein; CDC/NIAIDVAC-1; multivalent; malaria; vaccine;  
KW T-cell epitope; tetanus toxoid; antigenic epitope; treatment;  
KW circumsporozoite protein; CSP; sporozoite surface protein-2; SSP-2;  
KW liver stage antigen-1; LSA-1; merozoite surface protein-1; MSP-2;  
KW apical membrane antigen-1; AMA-1; erythrocyte binding antigen-175;  
KW EBA-175; rhoptry associated protein-1; RAP-1; gamete specific antigen;  
KW Pfg27; antiparasitic; prevention; anti-CDC/NIAIDVAC-1 antibody.  
XX  
OS Plasmodium falciparum.  
XX  
PN WC200011179-A1.  
XX  
PD 02-MAR-2000.  
XX  
PF 19-AUG-1999; 99WO-US18869.  
XX  
PR 21-AUG-1998; 98US-0097703.  
XX  
PA (NAIN-) NAT INST IMMUNOLOGY.  
XX  
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI Lal AA, Shi YP, Hasnain SE;  
XX  
DR WPI; 2000-237654/20.

## Example 2; Page 10; 18pp; English.

PS The polypeptide is prepd. by genetic engineering of genes encoding  
 XX the P. falciparum circumsporozoite (CS) protein [Dame et al.,  
 CC Science 225 : 593 (1984)], and the influenza virus non-structural  
 CC protein 1 (NS1), [Baez et al., Nucleic Acids Research, 8 : 5845  
 CC (1980)]. The DNA encoding the 1st 81 AAs of the N-terminal of NS1  
 CC (NS1-81) is linked via a synthetic sequence to DNA encoding Region  
 CC I contg. flanking region less the 18 AA signal region, which in  
 CC turn is fused to DNA encoding Region II-contg. flanking region.  
 CC This CS fusion is designated RLfAuth. The pro residue separating  
 CC the Asp (at the C-terminal of the linker) from RLfAuth is an arti-  
 CC fact of a filled in BamHI site; the Gly separating Region I and  
 CC Region II-contg. CS flanking regions is an artifact of a synthetic  
 CC FokI/ThIII I linker. The peptide can be used in a vaccine for  
 CC protection against malaria.  
 CC The complete nucleotide and AA sequences are given in EP-304720,  
 CC filed May 1, 1990.  
 CC See also AARI2306-R12311 and AARI3175-R13179.

Sequence 319 AA;

Query Match 100.0%; Score 127; DB 12; Length 319;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-10;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPKDELDYENDIEKKICKMEKCS 23  
 |||||

DB 275 kpkdeloyendiekkickmekcs 297

## RESULT 6

AARI3177  
 ID AARI3177 standard; Protein; 327 AA.

XX AARI3177;

DT 29-AUG-1991 (first entry)

DE NS1\_81-RLfAuth + (NANP)2.

KW Immunogenic determinant; circumsporozoite; CS; vaccine; malaria;  
 KW hybrid; influenza virus; non-structural protein 1; fusion.

XX Plasmodium falciparum.

OS Influenza virus (A/PR/8/34/).

XX Key Location/Qualifiers

FT Region 1..81 /label= N-terminal of NS1  
 FT /note= "Influenza virus nonstructural protein 1"

FT Peptide 82..87

FT Region /label= synthetic linker

FT /label= artifact

FT /note= "see comments"

FT Region 89..193

FT /label= AAs 19-123 of CS protein

FT /note= "Region I contg. flanking region less

FT signal sequence"

FT Region 194..201

FT /label= immunodominant repeat region

FT /note= "Two tetrapeptide repeat units"

FT Region 202

FT /label= artifact

FT /note= "see comments"

FT Region 203..327

FT /label= AAs 288-412 of CS protein

FT /note= "Region II flanking region"

XX EP432965-A.

XX

PD 19-JUN-1991.

XX 06-DEC-1990; 90EP-0313257.

XX 08-DEC-1989; 89US-0447746.

XX (SMIK ) SMITHKLINE BEECHAM.

XX (USSA ) US SEC OF THE ARMY.

XX (BIOM-) BIOMEDICAL RES INST.

XX Gross MS, Gordon DM, Hollingdale MR;

XX WPI; 1991-179771/25.

XX Polypeptide comprising immunogenic determinants from P falciparum  
 XX - for vaccine against malaria infection in humans.

XX Example 3; Page 10; 18pp; English.

XX The polypeptide is prepd. by genetic engineering of genes encoding  
 CC the P. falciparum circumsporozoite (CS) protein [Dame et al.,  
 CC Science 225 : 593 (1984)], and the influenza virus non-structural  
 CC protein 1 (NS1), [Baez et al., Nucleic Acids Research, 8 : 5845  
 CC (1980)]. The DNA encoding the 1st 81 AAs of the N-terminal of NS1  
 CC (NS1-81) is linked via a synthetic sequence to DNA encoding Region  
 CC I contg. flanking region less the 18 AA signal region. This is  
 CC linked to a synthetic sequence encoding two repeat units from the  
 CC immunodominant region, which in turn is fused to DNA encoding the  
 CC Region II-contg. flanking region. The pro residue separating the  
 CC Asp (at the C-terminal of the linker) from the Region I-contg. CS  
 CC flanking region is an artifact of a filled-in BamHI site; the Gly  
 CC separating the repeat units and the Region II-contg. CS flanking  
 CC region is an artifact of a synthetic FokI/ThIII I linker. The  
 CC peptide can be used in a vaccine for protection against malaria.  
 CC See also AARI2306-R12311 and AARI3175-R13179.

Sequence 327 AA;

Query Match 100.0%; Score 127; DB 12; Length 327;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-10;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPKDELDYENDIEKKICKMEKCS 23  
 |||||

DB 283 kpkdeloyendiekkickmekcs 305

## RESULT 7

AARI3178  
 ID AARI3178 standard; Protein; 335 AA.

XX AARI3178;

XX 29-AUG-1991 (first entry)

XX NS1\_81(NANP)4RLfAuth.

XX Immunogenic determinant; circumsporozoite; CS; vaccine; malaria;

XX hybrid; influenza virus; non-structural protein 1; fusion.

XX Plasmodium falciparum.

OS Influenza virus (A/PR/8/34).

XX Key Location/Qualifiers

FT Region 1..81 /label= N-terminal of NS1

FT /note= "Influenza virus nonstructural protein 1"

FT Region 82..97

FT /label= immunodominant repeat region

FT /note= "four tetrapeptide repeat units"

FT Peptide 98..103

FT /label= synthetic linker

XX PS Example 1; Page 7; 18pp; English.

XX CC The polypeptide is prep'd. by genetic engineering of genes encoding  
 CC the P. falciparum circumsporozoite (CS) protein [Dame et al.,  
 CC Science 225 : 593 (1984)], and the influenza-virus non-structural  
 CC protein 1 (NS1), [Baez et al., Nucleic Acids Research, 8 : 5845  
 CC (1980)]. The DNA encoding the 1st 81 AAs of the N-terminal of NS1  
 CC (NS1-81) is linked via a synthetic sequence to DNA encoding Region  
 CC I contg. flanking regionless the 18 AA signal region, which in  
 CC turn is fused to DNA encoding Region II-contg. flanking region  
 CC less the first nine N-terminal AAs. This CS fusion is designated  
 CC RLdelta9. The Pro residue separating the Asp (at the C-terminal  
 CC of the linker) from RLdelta9 is an artifact of a filled in BamHI  
 CC site. The peptide can be used in a vaccine for protection against  
 CC malaria.

XX CC See also AAR12306-RI2311 and AAR13176-RI3179.

XX SQ Sequence 309 AA;

Query Match 100.0%; Score 127; DB 12; Length 309;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-10;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPKDELDYENDIEKKICKMEKCS 23  
 |||||  
 Db 265 kpkdelyendiekkickmekcs 287

RESULT 4

AAR07945  
 ID AAR07945 standard; protein; 319 AA.

XX AC AAR07945;

XX DT 22-FEB-1991 (first entry)

XX DE NS181RLFAuth plasmid product.

XX KW Malaria; vaccine.

XX OS Plasmodium falciparum.

XX FH Key Location/Qualifiers

FT Domain 1..81

FT /label= NS181 protein fragment

FT /note= "from plasmid pMG-1"

FT Domain 89..193

FT /label= Fragment of circumsporozoite protein

FT Domain 204..319

FT /label= Fragment of circumsporozoite protein

XX EP398540-A.

XX PD 22-NOV-1990.

XX PF 01-MAY-1990; 90EP-0304720.

XX PR 03-MAY-1989; 89US-0346863.

XX PA (SMIK ) SMITHKLINE BEECHAM.

XX PI Gross MS, Young JF;

XX WPI: 1990-350299/47.

XX N-PSDB: AAQ06580.

XX New polypeptide used in malaria vaccine - comprises immunogenic  
 PT determinants from 1st and 2nd flanking regions of plasmodium  
 PT surface protein and intermediate repeat domain

XX Example 2; Page 11-12; 24pp; English.

XX CC The product is useful in preparation of vaccines for treatment and  
 CC prophylaxis of plasmodium sporozoite infection. It may be easily  
 CC produced in large pure quantities from a transformed E.coli  
 CC expression system.

XX SQ Sequence 319 AA;

Query Match 100.0%; Score 127; DB 11; Length 319;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPKDELDYENDIEKKICKMEKCS 23  
 |||||

Db 275 kpkdelyendiekkickmekcs 297

RESULT 5

AAR13176

ID AAR13176 standard; Protein; 319 AA.

XX AC AAR13176;

XX DT 29-AUG-1991 (first entry)

XX DE NS1\_81-RLFAuth.

XX KW Immunogenic determinant; circumsporozoite; CS; vaccine; malaria;  
 KW hybrid; Influenza virus; non-structural protein I; fusion.

XX OS Plasmodium falciparum.

XX OS Influenza virus (A/PR/8/34).

XX FH Key Location/Qualifiers

FT Region 1..81

FT /label= N-terminal of NS1

FT /note= "Influenza virus nonstructural protein 1"

FT Peptide 82..87

FT /label= synthetic linker

FT Region 88

FT /label= artifact

FT /note= "see comments"

FT Region 89..193

FT /label= AAs 19-123 of CS protein

FT /note= "Region I contg. flanking region less

FT signal sequence"

FT Region 194

FT /label= artifact

FT /note= "see comments"

FT Region 195..319

FT /label= AAs 288-412 of CS protein

FT /note= "Region II flanking region"

XX EP432965-A.

XX PD 19-JUN-1991.

XX PF 06-DEC-1990; 90EP-0313257.

XX PR 08-DEC-1989; 89US-0447746.

XX PA (SMIK ) SMITHKLINE BEECHAM.

XX PA (USSA ) US SEC OF THE ARMY.

XX PA (BIOM-) BIOMEDICAL RES INST.

XX PI Gross MS, Gordon DM, Hollingdale MR;

XX WPI: 1991-179771/25.

XX Polypeptide comprising immunogenic determinants-from P falciparum  
 PT - for vaccine against malaria infection in humans.



RC STRAIN-C57BL/6J; TISSUE-EMBRYONIC HEAD;  
RX MEDLINE=1085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Atakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
RA Arakawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schiraldi L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Guzman S., Hill D., Hofmann M., Hume D.A., Kaniya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,  
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
DR EMBL: AK014093; BAB39151.1; -  
DR MGD: MGI:1920407; 3110030A04Rik.  
DR InterPro: IPR002086; Aldenhydr\_dehydr.  
DR Pfam: PF00640; PID: 1.  
DR SMART: SM00462; PTB: 1.  
DR PROSITE: PS00687; ALDENHYDE\_DEHYDR\_GLU; UNKNOWN\_1.  
DR PROSITE: PS01179; PID: 1.  
FT NON\_TER 1  
SQ SEQUENCE 302 AA; 34207 MW; 8CB11440F898C65A CRC64;

Query Match 58.1%; Score 43; DB 11; Length 302;  
Best Local Similarity 57.1%; Pred. No. 11;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
QY 2 YIKANSKFIGITEL 15  
11 1:11:111:  
Db 22 YIPYNAKFLGSTEV 35

RESULT 10  
Q9CYD2 PRELIMINARY; PRT; 309 AA.  
AC Q9CYD2;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE 5730529006RIK PROTEIN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN 11  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE-EMBRYO;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Atakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
RA Arakawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schiraldi L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustingich S., Hill D., Hofmann M., Hume D.A., Kaniya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,  
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
DR EMBL: AK017798; BAB30939.1; -  
DR MGD: MGI:191926; 5730529006RIK.  
DR InterPro: IPR002086; Aldenhydr\_dehydr.  
DR Pfam: PF00640; PID: 1.  
DR SMART: SM00462; PTB: 1.  
DR PROSITE: PS00687; ALDENHYDE\_DEHYDR\_GLU; UNKNOWN\_1.  
DR PROSITE: PS01179; PID: 1.  
SQ SEQUENCE 309 AA; 35272 MW; 535D8EF33C0F406 CRC64;

Query Match 58.1%; Score 43; DB 11; Length 309;  
Best Local Similarity 57.1%; Pred. No. 12;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
QY 2 YIKANSKFIGITEL 15  
11 1:11:111:  
Db 24 YIPYNAKFLGSTEV 37

RESULT 11  
Q9YTK4 PRELIMINARY; PRT; 899 AA.  
AC Q9YTK4;  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
DE ORF 63.  
OS Ateline herpesvirus 3.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Gammaherpesvirinae; Rhadinovirus.  
OX NCBI\_TaxID=85618;  
RN 11  
RP SEQUENCE FROM N.A.  
RC STRAIN=73;  
RA Albrecht J.-C., Fleckenstein B.;  
RT "Primary structure of the Herpesvirus Ateles genome.";  
RL Submitted (AUG-1998) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AF083424; AAC95587.1; -  
SQ SEQUENCE 899 AA; 103389 MW; 19440A7944DE2531 CRC64;

Query Match 58.1%; Score 43; DB 12; Length 899;  
Best Local Similarity 64.3%; Pred. No. 36;  
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 1 OYIKANSKFIGITTE 14  
11 1:11:111:  
Db 124 OYITNSSTFGOTE 137

RESULT 12  
Q25961 PRELIMINARY; PRT; 1087 AA.  
AC Q25961;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE MAJOR SURFACE ANTIGEN (FRAGMENT).  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN 11  
RP SEQUENCE FROM N.A.  
RC STRAIN=RO-71;  
RX MEDLINE=92275047; PubMed=1592091;

SQ SEQUENCE 67 AA; 7449 MW; 42771AEDBABA6626 CRC64;

## Query Match

Best Local Similarity 58.8%; Score 43.5; DB 7; Length 67;  
Matches 9; Conservative 4; Mismatches 2; Indels 3; Gaps 1;

OY 1 QYKANS---KFIGITEL 15  
: : : : :  
Db 16 EYVRFNSTVGKFIGYTTEL 33

## RESULT 6

09CF66 PRELIMINARY; PRT; 180 AA.  
AC 09CF66;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE SPERMIDINE ACETYLTTRANSFERASE (EC 2.3.1.57).  
GN YQPF.  
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OX NCBI\_TaxID=1360;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=IL1403;  
RA Bolotin A., Wincker P., Manger S., Tallon O., Malarne K.,  
RT Weisenbach J., Ehrlich S.D., Sorokin A.;  
RT "The complete genome sequence of the lactic acid bacterium Lactococcus  
RT lactis.";  
RL Genome Res. 0:0-0(2001).  
DR EMBL: AE006391; AAK05713.1; -  
DR InterPro: IPR000182; Acetyltransf\_GCN5.  
DR Pfam: PF00583; Acetyltransf. 1.  
KW Transferase; Complete proteome.  
SQ SEQUENCE 180 AA; 21022 MW; 6DBD148524C0DF3C CRC64;

Query Match 58.1%; Score 43; DB 2; Length 180;  
Best Local Similarity 69.2%; Pred. No. 6.7;  
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 3 IKANSKFIGITEL 15  
: : : : :  
Db 65 IEANDFIGIVEL 77

## RESULT 7

09MCL7 PRELIMINARY; PRT; 250 AA.  
AC 09MCL7;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE ORF13.  
GN ORF13.  
OS Streptococcus thermophilus bacteriophage 7201.  
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.  
OX NCBI\_TaxID=112023;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20088830; PubMed=10620678;  
RA Stanley E., Walsh L., van der Zwet A., Fitzgerald G.F.,  
RT van Sinderen D.;  
RT "Identification of four loci isolated from two Streptococcus  
RT thermophilus phage genomes responsible for mediating bacteriophage  
RT resistance.";  
RL FEMS Microbiol. Lett. 182:271-277(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Stanley E., Walsh L., Fitzgerald G.F., van Sinderen D.;

RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF145054; AAF43506.1; -  
DR InterPro: IPR002295; D2IN6\_mlfase.  
DR InterPro: IPR001091; N4\_Mtase.  
DR InterPro: IPR002941; N6\_Mtase.  
DR Pfam: PF01555; N6\_Mtase; 1.  
DR PRINTS: PR00506; D2IN6MTFRASE.  
DR PRINTS: PR00508; S2IN4MTFRASE.  
SQ SEQUENCE 250 AA; 28744 MW; 725EA3B2B5B0D7F CRC64;

Query Match 58.1%; Score 43; DB 9; Length 250;  
Best Local Similarity 57.1%; Pred. No. 9.5;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 QYKANSKFIGITE 14  
: : : : :  
Db 131 QVLANMKIVGATE 144

## RESULT 8

09XJEB PRELIMINARY; PRT; 252 AA.  
AC 09XJEB;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE PUTATIVE TYPE II DNA METHYLTTRANSFERASE.  
OS Lactococcus lactis bacteriophage Iuc2009.  
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.  
OX NCBI\_TaxID=35241;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99240405; PubMed=10223975;  
RA McGirth S., Seegers J.F., Fitzgerald G.F., van Sinderen D.;  
RT "Molecular characterization of a phage-encoded resistance system in  
RT lactococcus lactis.";  
RL Appl. Environ. Microbiol. 65:1891-1899(1999).  
DR EMBL: AF109874; AAD37103.1; -  
DR InterPro: IPR001091; N4\_Mtase.  
DR InterPro: IPR002295; D2IN6\_mlfase.  
DR InterPro: IPR002941; N6\_Mtase.  
DR Pfam: PF01555; N6\_Mtase; 1.  
DR PRINTS: PR00506; D2IN6MTFRASE.  
DR PRINTS: PR00508; S2IN4MTFRASE.  
KW Transferase; Methyltransferase.  
SQ SEQUENCE 252 AA; 29357 MW; 4D5A59FF47363948 CRC64;

Query Match 58.1%; Score 43; DB 9; Length 252;  
Best Local Similarity 57.1%; Pred. No. 9.5;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 QYKANSKFIGITE 14  
: : : : :  
Db 130 QVLANMKIVGATE 143

## RESULT 9

09CRV4 PRELIMINARY; PRT; 302 AA.  
AC 09CRV4;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE 3110030A04RIK PROTEIN (FRAGMENT).  
GN 3110030A04RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.

Query Match 100.0%; Score 127; DB 9; Length 411;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-10;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPRDELVDYNDIEKKICKMEKCS 23  
 DB 367 kpxdeldyendlekkickmekcs 389

# RESULT 11

AAP60416  
 ID AAP60416 standard; Protein; 412 AA.

AC AAP60416;

DT 13-JUN-1991 (first entry)

DE CS protein of malaria parasite.

XX Sporozoite; vaccination.

OS Plasmodium falciparum.

Key Location/Qualifiers  
 FT Region 124..127  
 FT /label= Repeat unit

PN EPI66410-A.

PD 02-JAN-1986.

PF 24-JUN-1985; 85EP-0107794.

PR 26-JUN-1984; 84US-0624564.

PA (USDC ) US SEC OF COMMERCE.

PA (USGO ) US GOVERNMENT.

PA (USCA ) US SEC OF THE ARMY.

PI McCutchan TF, Dame JB, Williams JL, Schneider I;

DR MPI; 1986-008635/02.

DR N-PSDB; AAN60362.

PT New immunologically active pure synthetic peptide(s) - used for

PT protection against infection by malaria parasite.

PS Disclosure: Fig 2; 49pp; English.

XX The Plasmodium CS gene was used to isolate peptides capable of

CC inducing an immune response to the parasite. Peptide antigens may

CC be synthesised in pure form and used to generate an immune

CC response in vaccination against malaria. The featured repeat

CC units are claimed and must be present in copies of 2-1000.

XX Sequence 412 AA;

SO

Query Match 100.0%; Score 127; DB 7; Length 412;

Best Local Similarity 100.0%; Pred. No. 2.1e-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPRDELVDYNDIEKKICKMEKCS 23

DB 368 kpxdeldyendlekkickmekcs 390

XX 27-SEP-1993 (first entry)

XX RTS protein.

XX RTS; expression cassette; hybrid protein; S. cerevisiae; TDH3;

KW cloning; circumsporozoite protein; CSP; Plasmodium falciparum;

KW strain 7G8; hepatitis B virus; HBV; adv serotype; pres2 protein;

XX S protein.

XX Synthetic.

OS

Key Location/Qualifiers

FT Region 1

FT /note= "Derived from S. cerevisiae TDH3 gene sequence"

FT Region 2..4

FT /note= "Cloning artefact"

FT Protein 5..193

FT /note= "Represents amino acids 210-398 of the CSP of

FT Region 194..197

FT /note= "Carboxy terminal amino acids from HBV (adv

FT Protein 198..424

FT /note= "S protein of HBV (adv serotype)"

PN WO9310152-A.

PD 27-MAY-1993.

PF 11-NOV-1992; 92WO-EP02591.

PR 16-NOV-1991; 91GB-0024390.

PR 27-FEB-1992; 92US-0842694.

XX (SMK ) SMITHKLINE BEECHAM BIOLOGICALS.

XX Cohen J, De Wilde M;

XX MPI; 1993-182494/22.

XX N-PSDB; AAQ42566.

PT Hybrid protein comprising Plasmodium circumsporozoite protein and

PT Hbsag - useful as a vaccine for treating patients susceptible to

PT Plasmodium infections

PS Disclosure: Fig 5; 59pp; English.

XX This sequence represents the RTS hybrid protein which is encoded by

CC the RTS expression cassette. This hybrid consists of a methionine

CC residue derived from S. cerevisiae TDH3 gene sequence, three amino

CC acids, Met-Ala-Pro, derived from a nucleotide sequence created by

CC the cloning procedure used to construct the hybrid gene, a stretch

CC of 189 amino acids representing amino acids 210 to 398 of the

CC circumsporozoite protein (CSP) of Plasmodium falciparum strain 7G8,

CC an amino acid Arg created by the cloning procedure, four amino acids,

CC pro-Val-Thr-Asn, representing the four carboxy terminal residues of

CC hepatitis B virus (HBV), adv serotype, pres2 protein, and a stretch

CC of 226 amino acids specifying the S protein of HBV, adv serotype.

CC This protein, and RTS\* (see also AAR37797), may be combined with an

CC adjuvant and used in a vaccine for preventing plasmodium infections...

CC The vaccines produce a humoral response and also a cellular immune

CC response.

XX Sequence 424 AA;

SO

Query Match 100.0%; Score 127; DB 14; Length 424;

Best Local Similarity 100.0%; Pred. No. 2.2e-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPRDELVDYNDIEKKICKMEKCS 23

DB 368 kpxdeldyendlekkickmekcs 390

DB 163 kpkdeldyendiekkickmekes 185

# RESULT 13

ID AAP80835 standard; protein; 412 AA.

XX AAP80835;

DT 18-SEP-1990 (first entry)

DE Sequence encoded by the circumsporozoite (CS) gene of Plasmodium falciparum in lambda mpf.

XX Circumsporozoite gene; Plasmodium falciparum; lambda mpf; vaccine;  
KM yeast glyceraldehyde-3P-dehydrogenase gene (TDH3) promoter;  
KM yeast ornithine carbamoyl transferase gene (ARG3); repeat region.

OS Plasmodium falciparum

XX Key Location/Qualifiers

FT Region 124..147 /note="repeat region, repeat unit=NNPNVDP"

FT Region 148..207 /note="repeat region, repeat unit=NNNP"

FT Region 212..287 /note="repeat region, repeat unit=NNNP"

XX WO8805817-A.

XX 11-SEP-1988.

XX 25-JAN-1988; 88WO-BE00002.

XX 30-JAN-1987; 87US-0008791.

XX (SMIK) SMITH KLINE RIT SA.

XX De Wilde M, Gathoye AM;

XX WPI; 1988-235171/33.

DR N-PSDB; AAN81781.

PT Expression of P. falciparum circumsporozoite protein by yeast -  
PT using recombinant DNA vector having coding sequence linked to  
PT expression control sequence

XX Example 1; Fig 2a; 44pp; English.

XX Plasmid WR201 was obtained from the Walter Reed Army Institute of  
CC Research, and results from insertion of a 2.3 kb EcoRI fragment from  
CC lambda mpf encoding the complete CS protein gene P. falciparum into  
CC vector pUC8. A recombinant DNA vector is claimed, which comprises a DNA  
CC sequence contg. the coding sequence of the CS P. falciparum operatively  
CC linked to an expression control sequence. Pfief. expression control  
CC sequences include the yeast glyceraldehyde-3P-dehydrogenase gene (TDH3)  
CC promoter and the yeast ornithine carbamoyl transferase gene (ARG3)  
CC transcription termination region. A suitable coding sequence comprises  
CC the 1215bp StuI-RsaI fragment of WR201 contg. the P. falciparum CS  
CC protein coding sequence, minus its first 50bp. Also claimed is a  
CC transformed host cell, a method of culturing the cell to produce CS, the  
CC protein, and a vaccine. The vector comprises a 192 bp Sau3A fragment  
CC coding for 16 tetrapeptide repeats of the P. falciparum CS protein  
CC derived from Sau3A digestion of a 1215 bp StuI-RsaI fragment of WR201  
CC containing the P. falciparum CS protein coding sequence minus  
CC approx. its first 50 bp, or two, three, four or more tandem copies of  
CC such 192 bp Sau3A fragment.

XX Sequence 412 AA;

Query Match 96.9%; Score 123  
Best-Local Similarity 95.7%; Pred. No.

Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPKDELDYENDIEKKICKMEKES 23

DB 368 kpkdeldyendiekkickmekes 390

# RESULT 14

ID AAR37797 standard; Protein; 424 AA.

XX AAR37797;

DT 27-SEP-1993 (first entry)

DE RTS\* protein.

XX RTS; expression cassette; hybrid protein; S. cerevisiae; TDH3;

KM cloning; circumsporozoite protein; CSP; Plasmodium falciparum;

KM strain 7G8; hepatitis B virus; HBV; adv serotype; pres2 protein;

XX S protein.

XX Synthetic.

FT Key Location/Qualifiers

FT Region 1 /note="Derived from S. cerevisiae TDH3 gene sequence"

FT Region 2..4 /note="Cloning artefact"

FT Protein 5..193 /note="Represents amino acids 210-398 of the CSP of P. falciparum"

FT Region 194..197 /note="Carboxy terminal amino acids from HBV (adv serotype) pres2 protein"

FT Protein 198..424 /note="S protein of HBV (adv serotype)"

XX WO9310152-A.

XX 27-MAY-1993.

XX 11-NOV-1992; 92WO-EP02591.

XX 16-NOV-1991; 91GB-0024390.

XX 27-FEB-1992; 92US-0842694.

XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Cohen J, De Wilde M;

XX WPI; 1993-182494/22.

DR N-PSDB; AAQ42567.

PT Hybrid protein comprising Plasmodium circumsporozoite protein and  
PT HbsAg - useful as a vaccine for treating patients susceptible to  
PT Plasmodium infections

PS Disclosure; Fig 9; 59pp; English.

XX This sequence represents the RTS\* hybrid protein which is encoded by  
CC the RTS\* expression cassette. This hybrid consists of a methionine  
CC residue derived from S. cerevisiae TDH3 gene sequence, three amino  
CC acids, Met-Ala-Pro, derived from a nucleotide sequence created by  
CC the cloning procedure used to construct the hybrid gene, a stretch  
CC of 189 amino acids representing amino acids 210 to 398 of the  
CC circumsporozoite protein (CSP) of Plasmodium falciparum strain NF54,  
CC an amino acid Arg created by the cloning procedure, four amino acids,  
CC Pro-Val-Thr-Asn, representing the four carboxy terminal residues of  
CC hepatitis B virus (HBV), adv serotype, pres2 protein, and a stretch  
CC of 226 amino acids specifying the S protein of HBV, adv serotype.  
CC This protein, and RTS (see also AAR37796), may be combined with an  
CC adjuvant and used in a vaccine for preventing Plasmodium infections.

Job time: 419 sec

CC The vaccines produce a humoral response and also a cellular immune  
CC response.  
XX  
SQ Sequence 424 AA;

Query Match 95.3%; Score 121; DB 14; Length 424;  
Best Local Similarity 95.7%; Pred. No. 1.6e-09;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 KPKDELDYENDIEKKICKMEKCS 23  
DB 159 kpkdeldyendiekkickmekcs 181

RESULT 15

AA07290  
ID AA07290 standard; protein; 180 AA.

AC AA07290;

DT 28-JAN-1991 (first entry)

DE Circumsporozoite analogue Falci-parum 4.

KW CS protein; plasmodium; malaria; vaccine.

OS Synthetic.

PN EP392820-A.

PD 17-OCT-1990.

PF 11-APR-1990; 90EP-0303907.

PR 11-APR-1989; 89US-0336288.

PA (CHIR-) CHIRON CORP.

PI Barr PJ, Bathurst IC, Gibson HL;

DR WPI; 1990-314486/42.

DR N-PSDB; AA06153.

PT Recombinant Plasmodium circumsporozoite analogues - lacking  
PT one or more repeat epitope(s) for use as a malaria sub-unit  
PT vaccine.

PS Claim 10; Fig 10; 22pp; English.

XX The protein was produced by expression of a synthetic gene. The  
CC analogue comprises AAs 68-123 of the native P. falciparum CS  
CC protein, followed by four repeat sequences (three "B"s, i.e. MANP)  
CC and one "A", i.e. NYDP), followed by AAs 289-392 of the native  
CC protein. Reduction of the immunological dominance of the repeats  
CC relative to the epitopes in the regions flanking the repeats  
CC enhances sporozoite neutralising activity.  
CC See also AA07287-91.

XX  
SQ Sequence 180 AA;

Query Match 94.5%; Score 120; DB 11; Length 180;  
Best Local Similarity 95.7%; Pred. No. 8.6e-10;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 KPKDELDYENDIEKKICKMEKCS 23  
DB 157 kpkdeldyendiekkickmekcs 179



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 10:24:02 ; Search time 133.18 Seconds  
(without alignments)  
3.886 Million cell updates/sec

Title: US-09-763-397A-6  
127

Perfect score: 127  
Sequence: 1 KPDELVDYENDIEKRIKMEKCS 23

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/2/1aa/5A-COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B-COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/5A-COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/5B-COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/5A-COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/5B-COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	127	100.0	66	2	US-08-455-625-35
2	127	100.0	66	4	US-08-455-685-35
3	127	100.0	66	4	US-08-060-988A-35
4	127	100.0	66	5	PCT-US94-05142-35
5	127	100.0	412	1	US-08-313-288B-18
6	127	100.0	423	2	US-08-760-797A-1
7	127	100.0	424	2	US-08-932-929B-1
8	121	95.3	424	2	US-08-760-797A-3
9	121	95.3	424	2	US-08-932-929B-3
10	85	66.9	32	6	517861-7
11	82	69.6	15	2	US-08-318-856A-75
12	75	59.1	15	2	US-08-318-856A-76
13	57	44.9	19	6	517861-18
14	56.5	44.5	478	3	US-08-155-888-2
15	53	41.7	21	1	US-08-186-266-6
16	53	41.7	21	1	US-08-446-692-48
17	53	41.7	21	2	US-08-446-692-48
18	53	41.7	21	2	US-09-100-409A-54
19	53	41.7	21	5	PCT-US95-02121-97
20	53	41.7	21	5	PCT-US95-13841-20
21	53	41.7	33	2	US-08-446-692-27
22	53	41.7	33	2	US-08-446-692-27
23	51	40.2	378	6	US-08-488-351A-27
24	50	39.4	9	2	517843-9
25	47	37.0	9	2	US-08-318-856A-20
26	47	37.0	9	2	US-08-318-856A-28
27	46.5	36.6	492	4	US-09-215-221-25
					US-08-845-258-21

28	46.5	36.6	492	4	US-08-990-571-21	Sequence 21, Appl
29	46.5	36.6	492	4	US-08-723-142A-21	Sequence 21, Appl
30	46.5	36.6	503	4	US-08-845-258-52	Sequence 52, Appl
31	46.5	36.6	503	4	US-08-990-571-52	Sequence 52, Appl
32	46	36.2	863	4	US-09-238-303-11	Sequence 11, Appl
33	45	35.4	8	2	US-08-318-856A-17	Sequence 17, Appl
34	45	35.4	8	5	PCT-US95-02121-17	Sequence 17, Appl
35	45	35.4	9	2	US-08-318-856A-27	Sequence 27, Appl
36	44	34.6	607	2	US-08-472-534-5	Sequence 5, Appl
37	44	34.6	714	4	US-08-472-534-3	Sequence 3, Appl
38	44	34.6	1400	1	US-08-080-255-7	Sequence 7, Appl
39	44	34.6	1400	3	US-08-465-713-7	Sequence 7, Appl
40	44	34.6	1400	5	PCT-US93-05857-7	Sequence 7, Appl
41	44	34.6	3969	4	US-08-061-376-5	Sequence 5, Appl
42	43.5	34.3	1264	1	US-07-789-915A-6	Sequence 6, Appl
43	43.5	34.3	1264	1	US-08-005-002C-6	Sequence 6, Appl
44	43.5	34.3	1264	1	US-08-487-203A-6	Sequence 6, Appl
45	43	33.9	16	2	US-08-817-933A-7	Sequence 7, Appl

#### ALIGNMENTS

RESULT 1  
US-08-455-625-35  
Sequence 35, Application US/08455625  
Patent No. 5932218  
GENERAL INFORMATION:  
APPLICANT: Berzofsky, Jay A.  
APPLICANT: Ahlers, Jeffrey D.  
APPLICANT: Pendleton, C. D.  
APPLICANT: Nara, Peter  
TITLE OF INVENTION: COMPOSITE SYNTHETIC PEPTIDE CONSTRUCT  
TITLE OF INVENTION: ELICITING NEUTRALIZING ANTIBODIES AND CYTOTOXIC T  
TITLE OF INVENTION: LYMPHOCYTES AGAINST HIV  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/455,625  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/060,988  
FILING DATE: 14-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Svensson, Leonard R.  
REGISTRATION NUMBER: 30330  
REFERENCE/DOCKET NUMBER: 1173-434P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-205-8000  
TELEFAX: 703-205-8050  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 66 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLCULE TYPE: peptide  
FRAGMENT TYPE: Internal  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..66

OTHER INFORMATION: /label= peptide  
OTHER INFORMATION: /note= "peptide from P. falciparum CS antigen"  
US-08-455-625-35

Query Match 100.0%; Score 127; DB 2; Length 66;  
Best Local Similarity 100.0%; Pred. No. 8,9e-12;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPKDELVDYNDIEKKICKMEKCS 23  
Db 44 KPKDELVDYNDIEKKICKMEKCS 66

RESULT 2  
US-08-455-685-35

Sequence 35, Application US/08455685  
Patent No. 6214347

GENERAL INFORMATION:  
APPLICANT: Berzofsky, Jay A.  
APPLICANT: Ahlers, Jeffrey D.  
APPLICANT: Pendleton, C. David  
APPLICANT: Nara, Peter  
APPLICANT: Shirai, Mutsunori  
TITLE OF INVENTION: MULTIDETERMINANT PEPTIDES THAT ELICIT  
TITLE OF INVENTION: HELPER T-LYMPHOCYTE, CYTOTOXIC T LYMPHOCYTE AND  
TITLE OF INVENTION: NEUTRALIZING ANTIBODY RESPONSES AGAINST HIV-1  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02110-2804

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/455,685  
FILING DATE: 31-MAY-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/060,988  
FILING DATE: 14-MAY-1993

APPLICATION NUMBER: 07/847,311  
FILING DATE: 06-MAR-1992  
APPLICATION NUMBER: 07/751,998  
FILING DATE: 29-AUG-1991  
APPLICATION NUMBER: 07/148,692  
FILING DATE: 26-JAN-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Beattie, Ingrid A.  
REGISTRATION NUMBER: P-42,306  
REFERENCE/DOCKET NUMBER: 08830/022003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154

INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 66 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-455-685-35

Query Match 100.0%; Score 127; DB 4; Length 66;  
Best Local Similarity 100.0%; Pred. No. 8,9e-12;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPKDELVDYNDIEKKICKMEKCS 23  
Db 44 KPKDELVDYNDIEKKICKMEKCS 66

RESULT 3  
US-08-060-988A-35

Sequence 35, Application US/08060988A  
Patent No. 6294322

GENERAL INFORMATION:  
APPLICANT: Berzofsky, Jay A.  
APPLICANT: Ahlers, Jeffrey D.  
APPLICANT: Pendleton, C. David  
APPLICANT: Nara, Peter  
APPLICANT: Shirai, Mutsunori  
TITLE OF INVENTION: MULTIDETERMINANT PEPTIDES  
TITLE OF INVENTION: THAT ELICIT  
TITLE OF INVENTION: HELPER T-LYMPHOCYTE, CYTOTOXIC T LYMPHOCYTE AND  
TITLE OF INVENTION: NEUTRALIZING ANTIBODY RESPONSES AGAINST HIV-1  
NUMBER OF SEQUENCES: 48  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02110-2804

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/060,988A  
FILING DATE: 14-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/847,311  
FILING DATE: 06-MAR-1992

APPLICATION NUMBER: 07/751,998  
FILING DATE: 29-AUG-1991  
APPLICATION NUMBER: 07/148,692  
FILING DATE: 26-JAN-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Beattie, Ingrid A.  
REGISTRATION NUMBER: P-42,306  
REFERENCE/DOCKET NUMBER: 08830/022001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154

INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 66 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-060-988A-35

Query Match 100.0%; Score 127; DB 4; Length 66;  
Best Local Similarity 100.0%; Pred. No. 8,9e-12;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPKDELVDYNDIEKKICKMEKCS 23  
Db 44 KPKDELVDYNDIEKKICKMEKCS 66

RESULT 4  
PCT-US94-05142-35

Sequence 35, Application PC/TUS9405142  
GENERAL INFORMATION:  
APPLICANT:



;; TITLE OF INVENTION: COMPOSITE SYNTHETIC PEPTIDE CONSTRUCT  
;; TITLE OF INVENTION: ELICITING NEUTRALIZING ANTIBODIES AND CYTOTOXIC T  
;; TITLE OF INVENTION: LYMPHOCYTES AGAINST HIV  
;; NUMBER OF SEQUENCES: 36  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
;; STREET: P.O. Box 747  
;; CITY: Falls Church  
;; STATE: Virginia  
;; COUNTRY: USA  
;; ZIP: 22040-0747  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US94/05142  
;; FILING DATE: 13-MAY-1994  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/060,988  
;; FILING DATE: 14-MAY-1993  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Svensson, Leonard R.  
;; REGISTRATION NUMBER: 30330  
;; REFERENCE/DOCKET NUMBER: 1173-434P  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 703-205-8000  
;; TELEFAX: 703-205-8050  
;; INFORMATION FOR SEQ ID NO: 35:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 66 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; FRAGMENT TYPE: internal  
;; FEATURE:  
;; NAME/KEY: Peptide  
;; LOCATION: 1..66  
;; OTHER INFORMATION: /label= peptide  
;; PCT-US94-05142-35

Query Match 100.0%; Score 127; DB 5; Length 66;  
Best Local Similarity 100.0%; Pred. No. 8.9e-12;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KPDELVDYNDIEKKICKMEKCS 23  
Db 44 KPDELVDYNDIEKKICKMEKCS 66

RESULT 5  
US-08-313-288B-18  
; Sequence 18, Application US/08313288B  
; Patent No. 5750502  
; GENERAL INFORMATION:  
; APPLICANT: Jessell, Thomas M. and Avihu Klar  
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A  
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/313,288B  
;; FILING DATE: January 5, 1995  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: White, John P.  
;; REGISTRATION NUMBER: 28,678  
;; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (212) 278-0400  
;; TELEFAX: (212) 391-0526  
;; TELEX:  
;; INFORMATION FOR SEQ ID NO: 18:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 412 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; US-08-313-288B-18

Query Match 100.0%; Score 127; DB 1; Length 412;  
Best Local Similarity 100.0%; Pred. No. 6.3e-11;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KPDELVDYNDIEKKICKMEKCS 23  
Db 368 KPDELVDYNDIEKKICKMEKCS 390

RESULT 6  
US-08-760-797A-1  
; Sequence 1, Application US/08760797A  
; Patent No. 5928902  
; GENERAL INFORMATION:  
; APPLICANT: De Wilde, Michel  
; APPLICANT: Cohen, Joseph  
; TITLE OF INVENTION: Hybrid Protein Between CS  
; TITLE OF INVENTION: from Plasmodium and HBsAg  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/760,797A  
; FILING DATE: 04-DEC-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/442,612  
; FILING DATE: 17-MAY-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Baumeister, Kirk  
; REGISTRATION NUMBER: 33,833  
; REFERENCE/DOCKET NUMBER: B45015-1C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-270-5096  
; TELEFAX: 610-270-5090  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 423 amino acids

TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-760-797A-1

Query Match 100.0%; Score 127; DB 2; Length 423;  
Best Local Similarity 100.0%; Pred. No. 6,5e-11;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPKDELVDYNDIEKKICKMEKCS 23  
DB 162 KPKDELVDYNDIEKKICKMEKCS 184

RESULT 7  
US-08-932-929B-1  
Sequence 1, Application US/08932929B  
Patent No. 6169171  
GENERAL INFORMATION:  
APPLICANT: De Wilde, Michel  
APPLICANT: Cohen, Joseph  
TITLE OF INVENTION: Hybrid Protein Between CS  
TITLE OF INVENTION: from Plasmodium and HBSAG  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/932,929B  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/760,797  
FILING DATE: 04-DEC-1996  
APPLICATION NUMBER: 08/442,612  
FILING DATE: 17-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Baumeister, Kirk  
REGISTRATION NUMBER: 33,833  
REFERENCE/DOCKET NUMBER: B45015-1FWC2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-5096  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 424 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-932-929B-1

Query Match 100.0%; Score 127; DB 4; Length 424;  
Best Local Similarity 100.0%; Pred. No. 6,5e-11;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPKDELVDYNDIEKKICKMEKCS 23  
DB 163 KPKDELVDYNDIEKKICKMEKCS 185

RESULT 8

US-08-760-797A-3  
Sequence 3, Application US/08760797A  
Patent No. 5928902

GENERAL INFORMATION:  
APPLICANT: De Wilde, Michel  
APPLICANT: Cohen, Joseph  
TITLE OF INVENTION: Hybrid Protein Between CS  
TITLE OF INVENTION: from Plasmodium and HBSAG  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/760,797A  
FILING DATE: 04-DEC-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/442,612  
FILING DATE: 17-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Baumeister, Kirk  
REGISTRATION NUMBER: 33,833  
REFERENCE/DOCKET NUMBER: B45015-1C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-5096  
TELEFAX: 610-270-5090  
TELEX:

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 424 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-760-797A-3

Query Match 95.3%; Score 121; DB 2; Length 424;  
Best Local Similarity 95.7%; Pred. No. 4,8e-10;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 KPKDELVDYNDIEKKICKMEKCS 23  
DB 159 KPKDELVDYNDIEKKICKMEKCS 181

RESULT 9  
US-08-932-929B-3  
Sequence 3, Application US/08932929B  
Patent No. 6169171  
GENERAL INFORMATION:

APPLICANT: De Wilde, Michel  
APPLICANT: Cohen, Joseph  
TITLE OF INVENTION: Hybrid Protein Between CS  
TITLE OF INVENTION: from Plasmodium and HBSAG  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/932,929B  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/760,797  
FILING DATE: 04-DEC-1996  
APPLICATION NUMBER: 08/442,612  
FILING DATE: 17-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Baumeister, Kirk  
REGISTRATION NUMBER: 33,833  
REFERENCE/DOCKET NUMBER: B45015-1FWC2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-5096  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 424 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-932-929B-3

Query Match 95.3%; Score 121; DB 4; Length 424;  
Best Local Similarity 95.7%; Pred. No. 4.8e-10;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 KRPDELVDYNDIEKKICKMEKCS 23  
DB 159 KRPDELVDYNDIEKKICKMEKCS 181

RESULT 10  
517861-7  
PATENT NO. 517861  
APPLICANT: VENGARA, ULISES; RUIZ, ANDRES; FERREIRA, ARTURO;  
NUSSENZWEIG, RUTH S.; NUSSENZWEIG, VICTOR N.  
TITLE OF INVENTION: CROSS-REACTIVE AND PROTECTIVE EPITOPES  
OF CIRCUMPOROZOITE PROTEINS  
NUMBER OF SEQUENCES: 18  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/370,241  
FILING DATE: 22-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 115,634  
FILING DATE: 26-OCT-1987  
APPLICATION NUMBER: 649,903  
FILING DATE: 12-SEP-1984  
SEQ ID NO: 7  
LENGTH: 32  
517861-7

Query Match 66.9%; Score 85; DB 6; Length 32;  
Best Local Similarity 94.1%; Pred. No. 4.8e-06;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 KRPDELVDYNDIEKKIC 17  
DB 16 KRPDELVDYNDIEKKIC 32

RESULT 11  
US-08-318-856A-75  
SEQUENCE 75, Application US/0831856A  
PATENT NO. 5972351  
GENERAL INFORMATION:

APPLICANT: Adrian V.S. Hill, et al.  
TITLE OF INVENTION: PLASMODIUM FALCIPARUM MHC CLASS I-  
TITLE OF INVENTION: RESTRICTED CTL EPITOPES DERIVED FROM PRE-ERYTHROCYTIC STAGE  
TITLE OF INVENTION: ANTIGENS (AS AMENDED)  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.  
STREET: 2033 K Street, N.W., Suite 800  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20006  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk, 3.5 inch, 1.44 mb  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 5.1+  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/318,856A  
FILING DATE: October 3, 1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 92 08 068.8  
FILING DATE: April 3, 1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 92 17 704.7  
FILING DATE: August 20, 1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/GB93/00711  
FILING DATE: April 5, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee Cheng  
REGISTRATION NUMBER: 40,949  
REFERENCE/DOCKET NUMBER: 263-PP1R15770S  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 721-8200  
TELEFAX: (202) 721-8250  
INFORMATION FOR SEQ ID NO: 75:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acid residues  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-318-856A-75

Query Match 64.6%; Score 82; DB 2; Length 15;  
Best Local Similarity 100.0%; Pred. No. 5.8e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 ENDIEKKICKMEKCS 23  
DB 1 ENDIEKKICKMEKCS 15

RESULT 12  
US-08-318-856A-76  
SEQUENCE 76, Application US/0831856A  
PATENT NO. 5972351  
GENERAL INFORMATION:  
APPLICANT: Adrian V.S. Hill, et al.  
TITLE OF INVENTION: PLASMODIUM FALCIPARUM MHC CLASS I-  
TITLE OF INVENTION: RESTRICTED CTL EPITOPES DERIVED FROM PRE-ERYTHROCYTIC STAGE  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.  
STREET: 2033 K Street, N.W., Suite 800  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20006  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk, 3.5 inch, 1.44 mb  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 5.1+  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/318,856A  
FILING DATE: October 3, 1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 92 08 068.8  
FILING DATE: April 3, 1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 92 17 704.7  
FILING DATE: August 20, 1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/GB93/00711  
FILING DATE: April 5, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee Cheng  
REGISTRATION NUMBER: 40,949  
REFERENCE/DOCKET NUMBER: 263-PIR1577US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 721-8200  
TELEFAX: (202) 721-8250  
INFORMATION FOR SEQ ID NO: 76:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acid residues  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-318-856A-76

Query Match 59.1%; Score 75; DB 2; Length 15;  
Best Local Similarity 93.3%; Pred. No. 5.9e-05;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 ELDYNDIEKKICKM 19  
DB 1 ELDYNDIEKKICKM 15

RESULT 13  
517861-18  
PATENT NO. 517861  
APPLICANT: VERGARA, ULISES;RUIZ, ANDRES;FERREIRA, ARTURO;  
NUSSENZWEIG, KUTH S.;NUSSENZWEIG, VICTOR N.  
TITLE OF INVENTION: CROSS-REACTIVE AND PROTECTIVE EPITOPES  
OF CIRCUMPOROZOITE PROTEINS  
NUMBER OF SEQUENCES: 18  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/370,241  
FILING DATE: 22-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 115,634  
FILING DATE: 26-OCT-1987  
APPLICATION NUMBER: 649,903  
FILING DATE: 12-SEP-1984  
SEQ ID NO:18:  
LENGTH: 19  
517861-18

Query Match 44.9%; Score 57; DB 6; Length 19;  
Best Local Similarity 91.7%; Pred. No. 0.031;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KPKDELIDYNDI 12  
DB 7 KPKDELIDYNDI 18

RESULT 14

US-08-155-888-2  
Sequence 2, Application US/08155888  
Patent No. 6066623  
GENERAL INFORMATION:  
APPLICANT: Hoffman, Stephen L.  
APPLICANT: Hedstrom, Richard C.  
APPLICANT: Sedegah, Martha  
TITLE OF INVENTION: POLYNUCLEOTIDE VACCINE PROTECTIVE  
TITLE OF INVENTION: AGAINST MALARIA, METHODS OF PROTECTION AND VECTOR FOR  
DELIVERING POLYNUCLEOTIDE VACCINES  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Naval Medical Res. & Dev. Cnd.  
STREET: Bldg. 1, T-12 8901 Wisconsin Ave.  
CITY: Bethesda  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20889-5606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/155,888  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Spevack, A. David  
REGISTRATION NUMBER: 24,743  
REFERENCE/DOCKET NUMBER: N.C. 75,851  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 295-1022  
TELEFAX: (202) 295-6759  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 478 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-155-888-2

Query Match 44.5%; Score 56.5; DB 3; Length 478;  
Best Local Similarity 52.2%; Pred. No. 1.1;  
Matches 12; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

QY 1 KPKDELIDYNDIEKKICKMEKCS 23  
DB 435 KOPENLLE-DIDIEKKMKCS 456

RESULT 15  
US-08-186-266-6  
Sequence 6, Application US/08186266  
Patent No. 5662907  
GENERAL INFORMATION:  
APPLICANT: KUBO, Ralph T.  
APPLICANT: GREY, Howard M.  
APPLICANT: SETTY, Alessandro  
APPLICANT: CELIS, Esteban  
TITLE OF INVENTION: INDUCTION OF ANTI-TUMOR CYTOTOXIC  
TITLE OF INVENTION: T LYMPHOCYTES IN HUMANS USING  
TITLE OF INVENTION: SYNTHETIC PEPTIDE EPITOPES  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend Kourile and Crew  
STREET: Stuart Street Tower, One Market Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: US  
ZIP: 94105-1493  
COMPUTER READABLE FORM:

```

: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/186,266
: FILING DATE: 25-JAN-1994
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/159,339
: FILING DATE: 29-NOV-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/103,396
: FILING DATE: 06-AUG-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/027,746
: FILING DATE: 05-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/926,666
: FILING DATE: 07-AUG-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Bastian, Kevin L.
: REGISTRATION NUMBER: 34,774
: REFERENCE/DOCKET NUMBER: 14137-50-4
: TELEPHONE: (415) 543-9600
: TELEFAX: (415) 543-5043
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 21 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: FEATURE:
: NAME/KEY: Peptide
: LOCATION: 1..21
: OTHER INFORMATION: /note="Plasmodium falciparum CS
: OTHER INFORMATION: protein at positions 378-398."
: US-08-186-266-6

```

```

Query Match 41.7% Score 53; DB 1; Length 21;
Best Local Similarity 84.6%; Pred. No. 0.13;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 11 DIEKKICKMEKCS 23
    ||||| |||||
Db 1 DIEKKIAKMEKAS 13

```

Search completed: January 29, 2002, 10:24:03  
 Job time: 509 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 10:26:36 ; Search time 144.96 Seconds  
(without alignments)

12.086 Million cell updates/sec

Title: US-09-763-397A-6

Perfect score: 127  
Sequence: 1 KPKELDYENDIEKICKMEKCS 23

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR-G8:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	127	100.0	412	1	OZQOAF
2	127	100.0	424	1	A54533
3	124	97.6	442	2	A54529
4	122	96.1	388	2	A3756
5	121	95.3	405	2	S05428
6	71	55.9	38	2	C60657
7	68	53.5	38	2	D60657
8	68	53.5	38	2	E60657
9	65	51.2	38	2	A38657
10	65	51.2	38	2	B60657
11	63	49.6	38	2	H60657
12	63	49.6	38	2	B38869
13	62	48.8	38	2	G60657
14	62	48.8	38	2	F60657
15	62	48.8	189	2	B29795
16	61.5	48.4	388	2	JC6164
17	60.5	47.6	378	1	OZQOAB
18	60.5	47.6	378	1	OZQOAL
19	60.5	47.6	398	1	OZQOAS
20	60.5	47.6	401	1	OZQOAC
21	60.5	47.6	419	1	OZQOAM
22	58.5	46.1	332	1	OZQOMB
23	58.5	46.1	348	1	OZQOMB
24	56.5	44.5	264	1	A44969
25	56.5	44.5	367	1	OZQOMY
26	53	41.7	177	2	S57043
27	51	40.2	367	2	A29319
28	51	40.2	367	2	A23068
29	51	40.2	378	1	OZQOAV

30	51	40.2	386	2	A48571	circumsporozoite p
31	51	40.2	387	2	D41156	circumsporozoite p
32	51	40.2	387	2	C41156	circumsporozoite p
33	51	40.2	395	2	A41156	circumsporozoite p
34	49.5	39.0	993	2	B64695	type I restriction
35	49	38.6	387	2	B82891	ferichrome transp
36	49	38.6	429	2	A54504	circumsporozoite p
37	49	38.6	485	2	A60610	circumsporozoite p
38	49	38.6	802	2	T32448	hypothetical prote
39	48	37.8	778	2	A64656	hypothetical prote
40	48	37.8	778	2	C71944	hypothetical prote
41	48	37.8	1714	2	E71609	Ser/Thr protein k1
42	47.5	37.4	264	2	F85927	hypothetical prote
43	47.5	37.4	264	2	H65058	hypothetical prote
44	47.5	37.4	598	2	C81281	glutamine-fructos
45	47	37.0	175	2	G71480	hypothetical prote

## ALIGNMENTS

RESULT 1  
OZQOAF  
circumsporozoite protein - malaria parasite (Plasmodium falciparum) (isolate INTW22)  
C:Species: Plasmodium falciparum  
C:Date: 15-Nov-1984 #sequence\_revision 15-Nov-1984 #text\_change 09-Jun-2000  
C:Accession: A03388  
R:Dame, J.B.; Williams, J.L.; McCutchan, T.F.; Weber, J.L.; Wiltz, R.A.; Hochmeyer, W  
Science 225, 593-599, 1984  
A:Title: Structure of the gene encoding the immunodominant surface antigen on the spo  
A:Reference number: A03388; MUID:84250215  
A:Accession: A03388  
A:Molecule type: DNA  
A:Residues: 1-412 <DAM>  
A:Cross-References: GB:K02194; NID:9160160; PIDN:AAA29524.1; PID:9160161  
A:Experimental source: clone 7G8  
C:Comment: Residues 1-16 are the probable signal sequence.  
C:Comment: There are 41 copies of a 4-residue repeating unit in the middle domain of  
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology  
F:336-390/Domain: thrombospondin type 1 repeat homology <THR1>

Query Match 100.0%; Score 127; DB 1; Length 412;  
Best Local Similarity 100.0%; Pred. No. 8e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPKELDYENDIEKICKMEKCS 23  
|||||  
Db 368 KPKELDYENDIEKICKMEKCS 390

RESULT 2  
A54533  
circumsporozoite protein - malaria parasite (Plasmodium falciparum) (strain T4, Thai  
C:Species: Plasmodium falciparum  
C:Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 09-Jun-2000  
C:Accession: A54533  
R:del Portillo, H.A.; Nussenzweig, R.S.; Enea, V.  
Mol. Biochem. Parasitol. 24, 289-294, 1987  
A:Title: Circumsporozoite gene of a Plasmodium falciparum strain from Thailand.  
A:Reference number: A54533; MUID:87315205  
A:Accession: A54533  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-424 <DEL>  
A:Cross-References: GB:M19752; NID:9160216; PIDN:AAA29555.1; PID:9160217  
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology  
F:348-402/Domain: thrombospondin type 1 repeat homology <THR1>

Query Match 100.0%; Score 127; DB 2; Length 424;  
Best Local Similarity 100.0%; Pred. No. 8.2e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPKDELVDYNDIEKKICKMEKCS 23  
|||||  
DB 380 KPKDELVDYNDIEKKICKMEKCS 402

RESULT 3  
A54529  
circumsporozoite protein - malaria parasite (Plasmodium falciparum) (strain Wellcome)

C:Species: Plasmodium falciparum  
C:Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 09-Jun-2000  
C:Accession: A54529  
R:Lockyer, M.J.; Schwarz, R.T.  
Mol. Biochem. Parasitol. 22, 101-108, 1987  
A:Title: Strain variation in the circumsporozoite protein gene of Plasmodium falciparum.  
A:Reference number: A54529; MUID:87115616  
A:Accession: A54529  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-442 <LOC>  
A:Cross-References: GB:M15505; NID:g160214; PIDN:AAA29554.1; PID:g160215  
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology  
C:Keywords: tandem repeat  
F:366-420/Domain: thrombospondin type 1 repeat homology <THRI>

Query Match 97.6%; Score 124; DB 2; Length 442;  
Best Local Similarity 95.7%; Pred. No. 2.2e-09;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPKDELVDYNDIEKKICKMEKCS 23  
|||||  
DB 398 KPKDELVDYNDIEKKICKMEKCS 420

RESULT 4

A39756  
circumsporozoite protein - Plasmodium reichenowi  
C:Species: Plasmodium reichenowi  
C:Date: 14-Feb-1992 #sequence\_revision 14-Feb-1992 #text\_change 20-Aug-1999  
C:Accession: A39756  
R:Lal, A.A.; Goldman, I.F.  
J. Biol. Chem. 266, 6686-6689, 1991  
A:Title: Circumsporozoite protein gene from Plasmodium reichenowi, a chimpanzee malaria  
A:Reference number: A39756; MUID:91201303  
A:Accession: A39756  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-388 <LOC>  
A:Cross-References: GB:M60972; NID:g160228; PIDN:AAA29561.1; PID:g160229  
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology  
F:312-366/Domain: thrombospondin type 1 repeat homology <THRI>

Query Match 96.1%; Score 122; DB 2; Length 388;  
Best Local Similarity 91.3%; Pred. No. 3.6e-09;  
Matches 21; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPKDELVDYNDIEKKICKMEKCS 23  
|||||  
DB 344 KPKDELVDYNDIEKKICKMEKCS 366

RESULT 5

S05428  
circumsporozoite protein - malaria parasite (Plasmodium falciparum) (isolate NF54)  
C:Species: Plasmodium falciparum  
C:Date: 07-Jun-1990 #sequence\_revision 07-Jun-1990 #text\_change 09-Jun-2000  
C:Accession: S05428; A45527; I60657  
R:Campbell, J.R.  
Nucleic Acids Res. 17, 5854, 1989  
A:Title: DNA sequence of the gene encoding a Plasmodium falciparum malaria candidate vac  
A:Reference number: S05428; MUID:89345189

A:Accession: S05428

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-405 <CAM>

A:Cross-References: EMBL:X15363

R:Caspers, P.; Gentz, R.; Matile, H.; Pink, J.R.; Sinigaglia, F.

Mol. Biochem. Parasitol. 35, 185-190, 1989

A:Title: The circumsporozoite protein gene from NF54, a Plasmodium falciparum isolate

A:Reference number: A45527; MUID:89364998

A:Accession: A45527

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-405 <CAS>

A:Cross-References: GB:M22982; GB:J04650; NID:g160168; PIDN:AAA29527.1; PID:g160169

R:Lockyer, M.J.; Marsh, K.; Newbold, C.I.

Mol. Biochem. Parasitol. 37, 275-280, 1989

A:Title: Wild isolates of Plasmodium falciparum show extensive polymorphism in T cell

A:Reference number: A60657; MUID:90114334

A:Accession: A60657

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 319-336,354-373 <LOC>

A:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology

F:329-383/Domain: thrombospondin type 1 repeat homology <THRI>

Query Match 95.3%; Score 121; DB 2; Length 405;  
Best Local Similarity 95.7%; Pred. No. 5.2e-09;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KPKDELVDYNDIEKKICKMEKCS 23  
|||||  
DB 361 KPKDELVDYNDIEKKICKMEKCS 383

RESULT 6

C60657  
circumsporozoite protein - malaria parasite (Plasmodium falciparum) (isolate 7G8) (fr  
C:Species: Plasmodium falciparum  
C:Date: 03-Jun-1993 #sequence\_revision 03-Jun-1993 #text\_change 09-Jun-2000  
C:Accession: C60657  
R:Lockyer, M.J.; Marsh, K.; Newbold, C.I.  
Mol. Biochem. Parasitol. 37, 275-280, 1989  
A:Title: Wild isolates of Plasmodium falciparum show extensive polymorphism in T cell  
A:Reference number: A60657; MUID:90114334  
A:Accession: C60657  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-38 <LOC>  
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology

Query Match 55.9%; Score 71; DB 2; Length 38;  
Best Local Similarity 100.0%; Pred. No. 0.0035;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPKDELVDYNDIE 13  
|||||  
DB 26 KPKDELVDYNDIE 38

RESULT 7

D60657  
circumsporozoite protein - malaria parasite (Plasmodium falciparum) (isolate Wel) (fr  
C:Species: Plasmodium falciparum  
C:Date: 03-Jun-1993 #sequence\_revision 03-Jun-1993 #text\_change 09-Jun-2000  
C:Accession: D60657  
R:Lockyer, M.J.; Marsh, K.; Newbold, C.I.  
Mol. Biochem. Parasitol. 37, 275-280, 1989  
A:Title: Wild isolates of Plasmodium falciparum show extensive polymorphism in T cell  
A:Reference number: A60657; MUID:90114334  
A:Accession: D60657  
A:Status: preliminary; not compared with conceptual translation



A:Molecule type: DNA  
A:Residues: 1-38 <LOC>  
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology

Query Match 53.5%; Score 68; DB 2; Length 38;  
Best Local Similarity 92.3%; Pred. No. 0.009;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPKDELVDYNDIE 13  
||||:|||||  
DB 26 KPKDQLVDYNDIE 38

RESULT 8  
E60657  
circumsporozoite protein - malaria parasite (Plasmodium falciparum) (isolate Itc2) (frag  
C:Species: Plasmodium falciparum  
C:Date: 03-Jun-1993 #sequence\_revision 03-Jun-1993 #text\_change 09-Jun-2000  
C:Accession: E60657  
R:Lockyer, M.J.; Marsh, K.; Newbold, C.I.  
Mol. Biochem. Parasitol. 37, 275-280, 1989  
A:Title: Wild isolates of Plasmodium falciparum show extensive polymorphism in T cell ef  
A:Reference number: A60657; MUID:90114334  
A:Accession: E60657  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-38 <LOC>  
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology

Query Match 53.5%; Score 68; DB 2; Length 38;  
Best Local Similarity 92.3%; Pred. No. 0.009;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPKDELVDYNDIE 13  
||||:|||||  
DB 26 KPKDQLVDYNDIE 38

RESULT 9  
A38869  
circumsporozoite protein - malaria parasite (Plasmodium falciparum) (isolate 427-5) (fra  
C:Species: Plasmodium falciparum  
C:Date: 03-Jun-1993 #sequence\_revision 03-Jun-1993 #text\_change 09-Jun-2000  
C:Accession: A38869  
R:Lockyer, M.J.; Marsh, K.; Newbold, C.I.  
Mol. Biochem. Parasitol. 37, 275-280, 1989  
A:Title: Wild isolates of Plasmodium falciparum show extensive polymorphism in T cell ef  
A:Reference number: A60657; MUID:90114334  
A:Accession: A38869  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-38 <LOC>  
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology

Query Match 51.2%; Score 65; DB 2; Length 38;  
Best Local Similarity 92.3%; Pred. No. 0.023;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 KPKDELVDYNDIE 13  
||||:|||||  
DB 26 KPKDELVDYNDIE 38

RESULT 10  
B60657  
circumsporozoite protein - malaria parasite (Plasmodium falciparum) (isolate T9-98) (fra  
C:Species: Plasmodium falciparum  
C:Date: 03-Jun-1993 #sequence\_revision 03-Jun-1993 #text\_change 09-Jun-2000  
C:Accession: B60657  
R:Lockyer, M.J.; Marsh, K.; Newbold, C.I.

Mol. Biochem. Parasitol. 37, 275-280, 1989  
A:Title: Wild isolates of Plasmodium falciparum show extensive polymorphism in T cell  
A:Reference number: A60657; MUID:90114334  
A:Accession: B60657  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-38 <LOC>  
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology

Query Match 51.2%; Score 65; DB 2; Length 38;  
Best Local Similarity 92.3%; Pred. No. 0.023;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 KPKDELVDYNDIE 13  
||||:|||||  
DB 26 KPKDELVDYNDIE 38

RESULT 11  
H60657  
circumsporozoite protein - malaria parasite (Plasmodium falciparum) (isolate 406-10)  
C:Species: Plasmodium falciparum  
C:Date: 03-Jun-1993 #sequence\_revision 03-Jun-1993 #text\_change 09-Jun-2000  
C:Accession: H60657  
R:Lockyer, M.J.; Marsh, K.; Newbold, C.I.  
Mol. Biochem. Parasitol. 37, 275-280, 1989  
A:Title: Wild isolates of Plasmodium falciparum show extensive polymorphism in T cell  
A:Reference number: A60657; MUID:90114334  
A:Accession: H60657  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-38 <LOC>  
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology

Query Match 49.6%; Score 63; DB 2; Length 38;  
Best Local Similarity 84.6%; Pred. No. 0.043;  
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPKDELVDYNDIE 13  
||||:|||||  
DB 26 KPKDQLVDYNDIE 38

RESULT 12  
B38869  
circumsporozoite protein - malaria parasite (Plasmodium falciparum) (isolate 366-2) (f  
C:Species: Plasmodium falciparum  
C:Date: 03-Jun-1993 #sequence\_revision 03-Jun-1993 #text\_change 09-Jun-2000  
C:Accession: B38869  
R:Lockyer, M.J.; Marsh, K.; Newbold, C.I.  
Mol. Biochem. Parasitol. 37, 275-280, 1989  
A:Title: Wild isolates of Plasmodium falciparum show extensive polymorphism in T cell  
A:Reference number: A60657; MUID:90114334  
A:Accession: B38869  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-38 <LOC>  
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology

Query Match 49.6%; Score 63; DB 2; Length 38;  
Best Local Similarity 84.6%; Pred. No. 0.043;  
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPKDELVDYNDIE 13  
||||:|||||  
DB 26 KPKDQLVDYNDIE 38

RESULT 13  
G60657

circumsporozoite protein - malaria parasite (Plasmodium falciparum) (isolate 366-8) (fr  
 C:Species: Plasmodium falciparum  
 C:Date: 03-Jun-1993 #sequence\_revision 03-Jun-1993 #text\_change 09-Jun-2000  
 C:Accession: G60657  
 R:Lockyer, M.J.; Marsh, K.; Newbold, C.I.  
 Mol. Biochem. Parasitol. 37, 275-280, 1989  
 A:Title: Wild isolates of Plasmodium falciparum show extensive polymorphism in T cell ep  
 A:Reference number: A60657; MUID:90114334  
 A:Accession: G60657  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 1-38 <LOC>  
 C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology

Query Match 48.8%; Score 62; DB 2; Length 38;  
 Best Local Similarity 84.6%; Pred. No. 0.059; 1; Mismatches 0; Gaps 0;  
 Matches 11; Conservative 1; Indels 0;  
 QY 1 KPKDELVDYNDIE 13  
 ||||:||||  
 DB 26 KPKDLDYNDIE 38

RESULT 14  
 F60657  
 circumsporozoite protein - malaria parasite (Plasmodium falciparum) (isolate 366-1) (fr  
 C:Species: Plasmodium falciparum  
 C:Date: 03-Jun-1993 #sequence\_revision 03-Jun-1993 #text\_change 09-Jun-2000  
 C:Accession: F60657  
 R:Lockyer, M.J.; Marsh, K.; Newbold, C.I.  
 Mol. Biochem. Parasitol. 37, 275-280, 1989  
 A:Title: Wild isolates of Plasmodium falciparum show extensive polymorphism in T cell ep  
 A:Reference number: A60657; MUID:90114334  
 A:Accession: F60657  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 1-38 <LOC>  
 C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology

Query Match 48.8%; Score 62; DB 2; Length 38;  
 Best Local Similarity 84.6%; Pred. No. 0.059; 1; Mismatches 0; Gaps 0;  
 Matches 11; Conservative 1; Indels 0;  
 QY 1 KPKDELVDYNDIE 13  
 ||||:||||  
 DB 26 KPKDLDYNDIE 38

RESULT 15  
 B29795  
 circumsporozoite protein - malaria parasite (Plasmodium falciparum) (strain LB5) (fragme  
 C:Species: Plasmodium falciparum  
 C:Date: 18-Oct-1989 #sequence\_revision 21-Jul-1995 #text\_change 09-Jun-2000  
 C:Accession: B29795; A60657  
 R:de la Cruz, V.F.; Lal, A.A.; McCutchan, T.F.  
 J. Biol. Chem. 262, 11935-11939, 1987  
 A:Title: Sequence variation in putative functional domains of the circumsporozoite prote  
 A:Reference number: A92609; MUID:87308186  
 A:Accession: B29795  
 A:Molecule type: DNA  
 A:Residues: 1-161 <DEL>  
 A:Cross-references: GB:M17802; GB:M17803; GB:M17806  
 R:Lockyer, M.J.; Marsh, K.; Newbold, C.I.  
 Mol. Biochem. Parasitol. 37, 275-280, 1989  
 A:Title: Wild isolates of Plasmodium falciparum show extensive polymorphism in T cell ep  
 A:Reference number: A60657; MUID:90114334  
 A:Accession: A60657  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 152-189 <LOC>  
 C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology

Query Match 48.8%; Score 62; DB 2; Length 189;  
 Best Local Similarity 84.6%; Pred. No. 0.26; 1; Mismatches 0; Gaps 0;  
 Matches 11; Conservative 1; Indels 0;  
 QY 1 KPKDELVDYNDIE 13  
 ||||:||||  
 DB 177 KPKDLDYNDIE 189

Search completed: January 29, 2002, 10:26:36  
 Job time: 647 sec

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 11:13:38 ; Search time 80.65 Seconds

(Without alignments)  
10.456 Million cell updates/sec

Title: US-09-763-397a-6

Perfect score: 127  
Sequence: 1 KPKDELVDYNDIEKKICKMEKCS 23

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	127	100.0	412	1	CSP_PLAFA
2	127	100.0	424	1	CSP_PLAFA
3	124	97.6	442	1	CSP_PLAFA
4	122	96.1	388	1	CSP_PLAFA
5	121	95.3	397	1	CSP_PLAFA
6	60.5	47.6	378	1	CSP_PLAFA
7	60.5	47.6	378	1	CSP_PLAFA
8	60.5	47.6	398	1	CSP_PLAFA
9	60.5	47.6	401	1	CSP_PLAFA
10	60.5	47.6	419	1	CSP_PLAFA
11	58.5	46.1	339	1	CSP_PLAFA
12	58.5	46.1	347	1	CSP_PLAFA
13	56.5	44.5	367	1	CSP_PLAFA
14	53	41.7	177	1	CSP_PLAFA
15	51	40.2	343	1	CSP_PLAFA
16	51	40.2	378	1	CSP_PLAFA
17	51	40.2	386	1	CSP_PLAFA
18	49	38.6	393	1	CSP_PLAFA
19	49	38.6	429	1	CSP_PLAFA
20	48	37.8	629	1	CSP_PLAFA
21	48	37.8	613	1	CSP_PLAFA
22	47	37.0	720	1	CSP_PLAFA
23	46.5	36.6	663	1	CSP_PLAFA
24	46	36.2	253	1	CSP_PLAFA
25	45.5	35.8	321	1	CSP_PLAFA
26	45	35.4	388	1	CSP_PLAFA
27	45	35.4	492	1	CSP_PLAFA
28	45	35.4	590	1	CSP_PLAFA
29	45	35.4	607	1	CSP_PLAFA
30	45	35.4	609	1	CSP_PLAFA
31	45	35.4	721	1	CSP_PLAFA
32	44.5	35.0	315	1	CSP_PLAFA
33	44	34.6	345	1	CSP_PLAFA

34	44	34.6	607	1	DNAX_STRPN
35	44	34.6	931	1	CC21_SCHPO
36	44	34.6	1018	1	H1RA_CHICK
37	44	34.6	1070	1	YHV4_YEAST
38	44	34.6	3866	1	HRX_MOUSE
39	44	34.6	3969	1	HRX_HUMAN
40	43.5	34.3	514	1	SYS_METMY
41	43.5	34.3	630	1	Y242_MYCE
42	43.5	34.3	1272	1	UBP2_YEAST
43	43.5	34.3	5179	1	MUC2_HUMAN
44	43	33.9	365	1	GLN2_DROME
45	43	33.9	372	1	RECA_STRAM

## ALIGNMENTS

RESULT	ID	SP	PLAFA	STANDARD	PRT	412 AA
1	CSP_PLAFA					
AC	P02893					
DT	21-JUL-1986					
DT	21-JUL-1986					
DT	20-AUG-2001					
DE	CIRCUMSPOROZOITE PROTEIN PRECURSOR (CS)					
OS	Plasmodium falciparum					
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium					
OX	NCBI_TaxID=56133					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=84250215; PubMed=6204383					
RA	Dane J.B., Williams J.L., McCutchan T.F., Weber J.L., Wirtz R.A.,					
RA	Hochmeyer W.T., Maloy W.L., Haynes J.D., Schneider I., Roberts D.,					
RA	Sanders G.S., Reddy E.P., Digs C.L., Miller L.H.;					
RT	Structure of the gene encoding the immunodominant surface antigen on					
RT	the sporozoite of the human malaria parasite Plasmodium falciparum.;					
RL	Science 225:593-599(1984).					
CC	-1- FUNCTION: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT					
CC	SURFACE ANTIGEN ON THE SPOROZOITE (THE INFECTIVE STAGE OF THE					
CC	MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE					
CC	VERTEBRATE HOST).					
CC	-1- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR					
CC	ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES					
CC	WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.					
CC	-1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.					
CC	-----					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (See <a href="http://www.isb-sdb.ch/announce/">http://www.isb-sdb.ch/announce/</a>					
CC	or send an email to <a href="mailto:license@sdb.ch">license@sdb.ch</a> ).					
CC	-----					
DR	EMBL; K02194; AAA29524.1; -					
DR	PIR; A03388; OZ0AF.					
DR	InterPro; IPR003067; Circmsprzoite.					
DR	InterPro; IPR000884; TSP1.					
DR	Pfam; PF00090; tsp_1; 1.					
DR	PRINTS; PR01303; CRCMSPRZOITE.					
DR	SMART; SM00209; TSP1; 1.					
KW	Malaria; Sporozoite; Repeat; Signal.					
FT	SIGNAL					
FT	CHAIN					
FT	DOMAIN					
SO	SEQUENCE					

Query Match 100.0% Score 127; DB 1; Length 412;  
Best Local Similarity 100.0% Pred. No. 5.9e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 368 KPDELVDYENDIEKKICKMEKCS 390

## RESULT 2

CSP\_PLAFT STANDARD: PRT: 424 AA.  
ID CSP\_PLAFT  
AC P13814:  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE CIRCUMSPOROZOITE PROTEIN PRECURSOR (CS).  
OS Plasmodium falciparum (isolate t4 / Thailand).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5846;  
[1]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=8715205; Pubmed=3306373;  
RA del Portillo H.A., Nussenzweig R.S., Enea V.;  
RT "Circumsporozoite gene of a Plasmodium falciparum strain from Thailand."  
RL Mol. Biochem. Parasitol. 24:289-294(1987).  
CC -1- FUNCTION: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT SURFACE ANTIGEN ON THE SPOOROZOITE (THE INFECTIVE STAGE OF THE MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE VERTEBRATE HOST).  
CC -1- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.  
CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M19752; AAA29555.1; -  
CC PIR: A54533; A54533.  
CC InterPro: IPR003067; Crcmsprzoite.  
CC InterPro: IPR000884; TSP1.  
CC Pfam: PF00090; tsp\_1; 1.  
CC PRINTS: PR01303; CRCMSPRZOITE.  
CC SMART: SMO0209; TSP1; 1.  
CC Malaria; Sporozoite; Repeat; Signal.  
CC SIGNAL 1 16 CIRCUMSPOROZOITE PROTEIN.  
CC CHAIN 17 424 45 X 4 AA TANDM REPEATS OF N-A-N-P.  
CC DOMAIN 123 300  
CC SEQUENCE 424 AA; 45610 MW; 710AB14238786CD9 CRC64;  
Query Match 100.0%; Score 127; DB 1; Length 424;  
Best Local Similarity 100.0%; Pred. No. 6, 1e-10;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPDELVDYENDIEKKICKMEKCS 23

Db 380 KPDELVDYENDIEKKICKMEKCS 402

RESULT 3  
CSP\_PLAFT STANDARD: PRT: 442 AA.  
ID CSP\_PLAFT  
AC P08307;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE CIRCUMSPOROZOITE PROTEIN PRECURSOR (CS).  
OS Plasmodium falciparum (isolate Wellcome).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5848;

[1]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=87115616; Pubmed=3543671;  
RA Lockyer M.J., Schwarz R.T.;  
RT "Strain variation in the circumsporozoite protein gene of Plasmodium falciparum."  
RL Mol. Biochem. Parasitol. 22:101-108(1987).  
CC -1- FUNCTION: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT SURFACE ANTIGEN ON THE SPOOROZOITE (THE INFECTIVE STAGE OF THE MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE VERTEBRATE HOST).  
CC -1- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.  
CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M15505; AAA29554.1; -  
CC PIR: A54529; A54529.  
CC InterPro: IPR003067; Crcmsprzoite.  
CC InterPro: IPR000884; TSP1.  
CC Pfam: PF00090; tsp\_1; 1.  
CC PRINTS: PR01303; CRCMSPRZOITE.  
CC SMART: SMO0209; TSP1; 1.  
CC Malaria; Sporozoite; Repeat; Signal.  
CC SIGNAL 1 16 CIRCUMSPOROZOITE PROTEIN.  
CC CHAIN 17 442 47 X 4 AA TANDM REPEATS OF N-A-N-P.  
CC DOMAIN 130 320  
CC SEQUENCE 442 AA; 47402 MW; BD57A9A152B85E03 CRC64;

Query Match 97.6%; Score 124; DB 1; Length 442;  
Best Local Similarity 95.7%; Pred. No. 1, 6e-09;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPDELVDYENDIEKKICKMEKCS 23

Db 398 KPDELVDYENDIEKKICKMEKCS 420

RESULT 4  
CSP\_PLAFT STANDARD: PRT: 388 AA.  
ID CSP\_PLAFT  
AC P26694;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE CIRCUMSPOROZOITE PROTEIN PRECURSOR (CS).  
OS Plasmodium reichenowi.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5854;  
[1]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=91201303; Pubmed=2016283;  
RA Lal A.A., Goldman I.F.;  
RT "Circumsporozoite protein gene from Plasmodium reichenowi, a chimpanzee malaria parasite evolutionarily related to the human malaria parasite Plasmodium falciparum."  
RL J. Biol. Chem. 266:6686-6689(1991).  
CC -1- FUNCTION: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT SURFACE ANTIGEN ON THE SPOOROZOITE (THE INFECTIVE STAGE OF THE MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE VERTEBRATE HOST).  
CC -1- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.

CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M60972; AAA29561.1; -  
DR PIR: A39756; A39756.  
DR InterPro: IPR003067; Circmsporzoite.  
DR Pfam: PF00090; TSP.1; 1.  
DR PRINTS: PR01303; Circmsporzoite.  
DR SMART: SM00209; TSP1.1  
KM Malaria; Sporozoite; Repeat; Signal.  
FT SIGNAL 1 16 PROBABLE.  
FT CHAIN 17 388 CIRCUMSPOROZOITE PROTEIN.  
FT DOMAIN 120 267 62 X 4 AA TANDEM REPEATS OF N-A-N-P.  
SQ SEQUENCE 388 AA; 42245 MW; C031EEFBE2E35604 CRC64;  
  
Query Match 96.1%; Score 122; DB 1; Length 388;  
Best Local Similarity 91.3%; Pred. No. 2.6e-09;  
Matches 21; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
Oy 1 KPKDELVDYNDIEKKICKMEKCS 23  
Db 344 KPKDQLDYNDIEKKICKMEKCS 366  
  
RESULT 5  
CSP\_PLAFO STANDARD: PRT: 397 AA.  
ID CSP\_PLAFO  
AC P19597; Q25798;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE CIRCUMSPOROZOITE PROTEIN PRECURSOR (CS).  
OS Plasmodium falciparum (isolate NF54).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5843;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89345189; PubMed=2668895;  
RA Campbell J.R.;  
RT "DNA sequence of the gene encoding a Plasmodium falciparum malaria  
RT candidate vaccine antigen.";  
RL Nucleic Acids Res. 17:5854-5854(1989).  
RN [2]  
RP REVISIONS.  
RA Campbell J.R.;  
RL Submitted (JUL-1997) to the EMBL/Genbank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92155298; PubMed=1346766;  
RA Davis J.R., Cortese J.F., Herrington D.A., Murphy J.R., Clyde D.F.,  
RA Thomas A.W., Bagar S., Cochran M.A., Thanassi J., Levine M.M.,  
RA Hackett C.S.;  
RT "Plasmodium falciparum: in vitro characterization and human  
RT infectivity of a cloned line.";  
RL Exp. Parasitol. 74:159-168(1992).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89364998; PubMed=2671723;  
RA Caspers P., Gentz R., Matile H., Pink J.R., Slingagla F.;  
RT "The circumsporozoite protein gene from NF54, a Plasmodium falciparum  
RT isolate used in malaria vaccine trials.";  
RL Mol. Biochem. Parasitol. 35:185-190(1989).  
CC -1- FUNCTION: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT  
CC SURFACE ANTIGEN ON THE SPOOROZOITE (THE INFECTIVE STAGE OF THE

CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE  
CC VERTEBRATE HOST).  
CC -1- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR  
CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES  
CC WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.  
CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X15363; CAA33421.1; -  
DR EMBL: M63886; AAA29521.1; -  
DR EMBL: M22982; AAA29527.1; -  
DR PIR: S05428; S05428.  
DR PIR: A45527; A45527.  
DR InterPro: IPR003067; Circmsporzoite.  
DR InterPro: IPR000884; TSP1.  
DR Pfam: PF00090; TSP.1; 1.  
DR PRINTS: PR01303; Circmsporzoite.  
DR SMART: SM00209; TSP1.1  
KM Malaria; Sporozoite; Repeat; Signal.  
FT SIGNAL 1 16 PROBABLE.  
FT CHAIN 17 397 CIRCUMSPOROZOITE PROTEIN.  
FT DOMAIN 101 272 43 X 4 AA TANDEM REPEATS OF N-A-N-P.  
FT CONFLICT 194 194 A -> ANPANA (IN REF. 4).  
SQ SEQUENCE 397 AA; 42646 MW; 9E81146F59BCEA3 CRC64;

Query Match 95.3%; Score 121; DB 1; Length 397;  
Best Local Similarity 95.7%; Pred. No. 3.6e-09;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Oy 1 KPKDELVDYNDIEKKICKMEKCS 23  
Db 353 KPKDELVDYNDIEKKICKMEKCS 375  
  
RESULT 6  
CSP\_PLACB STANDARD: PRT: 378 AA.  
ID CSP\_PLACB  
AC P08672;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE CIRCUMSPOROZOITE PROTEIN PRECURSOR (CS).  
OS Plasmodium cynomolgi (strain Berok).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5828;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87102878; PubMed=3802196;  
RA Galinski M.R., Arnot D.E., Cochran A.H., Barnwell J.W.,  
RA Nussenzweig R.S., Enea V.;  
RT "The circumsporozoite gene of the Plasmodium cynomolgi complex.";  
RL Cell 48:311-319(1987).  
CC -1- DOMAIN: THERE ARE 10 TANDEM COPIES OF A 9-RESIDUE REPEAT (PRECEDED  
CC BY A 6-RESIDUE INCOMPLETE REPEAT) AND 3 TANDEM COPIES OF A 16-  
CC RESIDUE REPEAT (FOLLOWED BY 3 SHORTER, INCOMPLETE COPIES).  
CC -1- MISCELLANEOUS: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT  
CC SURFACE ANTIGEN ON THE SPOOROZOITE (THE INFECTIVE STAGE OF THE  
CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE  
CC VERTEBRATE HOST).  
CC -1- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR  
CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES  
CC WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.  
CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 DR EMBL; M15104; AAA29532.1; -  
 DR PIR; D26255; OZ20AB.  
 DR InterPro; IPR003067; Crcmsprzoite.  
 DR InterPro; IPR000884; TSP1.  
 DR Pfam; PF00090; TSP1; 1.  
 DR PRINTS; PR01303; Crcmsprzoite.  
 DR SMART; SM00209; TSP1; 1.  
 KW Malaria; Sporozoite; Repeat; Signal.  
 FT SIGNAL 1 19 PROBABLE.  
 FT CHAIN 20 378 CIRCUMSPOROZOITE PROTEIN.  
 FT DOMAIN 97 192 10.5 x 9 AA REPEATS.  
 FT DOMAIN 193 240 3 x 16 AA TANDEM REPEATS.  
 FT REPEAT 241 251  
 FT REPEAT 252 260  
 FT REPEAT 261 268  
 SQ SEQUENCE 378 AA; 36286 MW; 779BA081C140793F CRC64;

Query Match 47.6%; Score 60.5; DB 1; Length 378;  
 Best Local Similarity 47.8%; Pred. No. 0.43;  
 Matches 11; Conservative 6; Mismatches 5; Indels 1; Gaps 1;  
 Oy 1 KPDELVDYENDIEKKICKMEKCS 23  
 Db 335 KPDELVDY-NDLETEVCTMDKCA 356.

RESULT 7  
 CSP\_PLACC STANDARD; PRT: 378 AA.  
 ID CSP\_PLACC  
 AC P08673;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE CIRCUMSPOROZOITE PROTEIN PRECURSOR (CS).  
 OS Plasmodium cynomolgi (strain London).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 ON NCBI\_TaxId=5831;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87102878; PubMed=3802196;  
 RA Galinski M.R., Arnot D.E., Cochrane A.H., Barnwell J.W.,  
 RA Nussenzweig R.S., Enea V.;  
 RT "The circumsporozoite gene of the Plasmodium cynomolgi complex.";  
 RL Cell 48:311-319(1987).  
 CC -1- MISCELLANEOUS: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT  
 CC SURFACE ANTIGEN ON THE SPOROZOITE (THE INFECTIVE STAGE OF THE  
 CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE  
 CC VERTEBRATE HOST).  
 CC -1- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR  
 CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES  
 CC WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.  
 CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR Pfam; PF00090; TSP1; 1.  
 DR PRINTS; PR01303; Crcmsprzoite.  
 DR SMART; SM00209; TSP1; 1.  
 KW Malaria; Sporozoite; Repeat; Signal.  
 FT SIGNAL 1 19 PROBABLE.  
 FT CHAIN 20 378 CIRCUMSPOROZOITE PROTEIN.  
 FT DOMAIN 98 211 18 x 6 AA TANDEM REPEATS OF D-G-A-R-A-  
 FT [EA].  
 FT DOMAIN 212 277 6 x 11 AA TANDEM REPEATS OF G-N-Q-A-G-G-  
 FT O-A-G-A-G.  
 SQ SEQUENCE 378 AA; 37462 MW; 8295A913C36420C5 CRC64;

Query Match 47.6%; Score 60.5; DB 1; Length 378;  
 Best Local Similarity 47.8%; Pred. No. 0.43;  
 Matches 11; Conservative 6; Mismatches 5; Indels 1; Gaps 1;  
 Oy 1 KPDELVDYENDIEKKICKMEKCS 23  
 Db 335 KPDELVDY-NDLETEVCTMDKCA 356

RESULT 8  
 CSP\_PLACC STANDARD; PRT: 398 AA.  
 ID CSP\_PLACC  
 AC P08673;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE CIRCUMSPOROZOITE PROTEIN PRECURSOR (CS).  
 OS Plasmodium cynomolgi (strain Ceylon).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 ON NCBI\_TaxId=5829;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87102878; PubMed=3802196;  
 RA Galinski M.R., Arnot D.E., Cochrane A.H., Barnwell J.W.,  
 RA Nussenzweig R.S., Enea V.;  
 RT "The circumsporozoite gene of the Plasmodium cynomolgi complex.";  
 RL Cell 48:311-319(1987).  
 CC -1- MISCELLANEOUS: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT  
 CC SURFACE ANTIGEN ON THE SPOROZOITE (THE INFECTIVE STAGE OF THE  
 CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE  
 CC VERTEBRATE HOST).  
 CC -1- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR  
 CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES  
 CC WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.  
 CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M15103; AAA29533.1; -  
 DR PIR; C26255; OZ20AS.  
 DR InterPro; IPR003067; Crcmsprzoite.  
 DR InterPro; IPR000884; TSP1.  
 DR Pfam; PF00090; TSP1; 1.  
 DR PRINTS; PR01303; Crcmsprzoite.  
 DR SMART; SM00209; TSP1; 1.  
 KW Malaria; Sporozoite; Repeat; Signal.  
 FT SIGNAL 1 19 PROBABLE.  
 FT CHAIN 20 398 CIRCUMSPOROZOITE PROTEIN.  
 FT DOMAIN 97 249 17 x 9 AA TANDEM REPEATS OF A-G-N-N-A-A-  
 FT A-G-E.  
 SQ SEQUENCE 398 AA; 37718 MW; 6DFA2E8A6ED05DF CRC64;

Query Match 47.6%; Score 60.5; DB 1; Length 398;



Best Local Similarity 47.8%, Pred. No. 0.45;  
Matches 11: Conservative 6; Mismatches 5; Indels 1; Gaps 1;  
Oy 1 KPKDELVDYNDIEKKICKMEKCS 23  
DB 355 KKPELD-ANDELEVCTMDKCA 376

RESULT 9  
CSP\_PLACG STANDARD: PRT: 401 AA.  
ID CSP\_PLACG STANDARD: PRT: 401 AA.  
AC P08674;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE CIRCUMSPOROZOITE PROTEIN PRECURSOR (CS).  
OS Plasmodium cynomolgi (strain Gomak).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5830;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87102878; PubMed=3802196;  
RA Galinski M.R., Arnot D.E., Cochran A.H., Barnwell J.W.,  
RA Nussenzweig R.S., Enea V.;  
RT "The circumsporozoite gene of the Plasmodium cynomolgi complex.";  
RL Cell 48:311-319(1987).  
CC -1- MISCELLANEOUS: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT  
CC SURFACE ANTIGEN ON THE SPOOROZOITE (THE INFECTIVE STAGE OF THE  
CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE  
CC VERTEBRATE HOST).  
CC -1- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR  
CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES  
CC WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.  
CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: M15100; AAA29536.1; -  
DR PIR: E26255; OZ2OAC.  
DR InterPro: IPR003067; Circmsprzoite.  
DR InterPro: IPR000884; TSP1.  
DR Pfam: PF00090; tsp\_1; 1.  
DR PRINTS: PR01303; CRCMSPRZOITE.  
DR SMART: SM00209; TSP1; 1.  
KW Malaria; Sporozoite; Repeat; Signal.  
FT SIGNAL 1 19 PROBABLE.  
FT CHAIN 20 401 CIRCUMSPOROZOITE PROTEIN.  
FT DOMAIN 98 278 17 X 11 AA TANDEM REPEATS OF [DG]-G-A-A-  
FT A-A-G-G-G-N.  
SO SEQUENCE 401 AA; 36664 MW; 57D666268238903E CRC64;

Query Match 47.6%, Score 60.5; DB 1; Length 401;  
Best Local Similarity 47.8%; Pred. No. 0.45;  
Matches 11: Conservative 6; Mismatches 5; Indels 1; Gaps 1;

Oy 1 KPKDELVDYNDIEKKICKMEKCS 23  
DB 358 KKPELD-ANDELEVCTMDKCA 379

RESULT 10  
CSP\_PLACG STANDARD: PRT: 419 AA.  
ID CSP\_PLACG STANDARD: PRT: 419 AA.  
AC P08676;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE CIRCUMSPOROZOITE PROTEIN PRECURSOR (CS).  
OS Plasmodium cynomolgi (strain Mulligan/NH).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5832;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87102878; PubMed=3802196;  
RA Galinski M.R., Arnot D.E., Cochran A.H., Barnwell J.W.,  
RA Nussenzweig R.S., Enea V.;  
RT "The circumsporozoite gene of the Plasmodium cynomolgi complex.";  
RL Cell 48:311-319(1987).  
CC -1- MISCELLANEOUS: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT  
CC SURFACE ANTIGEN ON THE SPOOROZOITE (THE INFECTIVE STAGE OF THE  
CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE  
CC VERTEBRATE HOST).  
CC -1- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR  
CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES  
CC WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.  
CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.  
-----

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL: M15102; AAA29539.1; -  
DR PIR: E26255; OZ2OAM.  
DR InterPro: IPR003067; Circmsprzoite.  
DR InterPro: IPR000884; TSP1.  
DR Pfam: PF00090; tsp\_1; 1.  
DR PRINTS: PR01303; CRCMSPRZOITE.  
DR SMART: SM00209; TSP1; 1.  
KW Malaria; Sporozoite; Repeat; Signal.  
FT SIGNAL 1 19 PROBABLE.  
FT CHAIN 20 419 CIRCUMSPOROZOITE PROTEIN.  
FT DOMAIN 99 314 54 X 4 AA TANDEM REPEATS OF N-A-[DG]-G.  
SO SEQUENCE 419 AA; 38924 MW; 8F46CDBA184EF4 CRC64;

Query Match 47.6%, Score 60.5; DB 1; Length 419;  
Best Local Similarity 47.8%; Pred. No. 0.47;  
Matches 11: Conservative 6; Mismatches 5; Indels 1; Gaps 1;

Oy 1 KPKDELVDYNDIEKKICKMEKCS 23  
DB 376 KKPELDV-NDLEVECTMDKCA 397

RESULT 11  
CSP\_PLACG STANDARD: PRT: 339 AA.  
ID CSP\_PLACG STANDARD: PRT: 339 AA.  
AC P06915;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE CIRCUMSPOROZOITE PROTEIN PRECURSOR (CS).  
OS Plasmodium berghei.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5821;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87089740; PubMed=24232395;  
RA Eichinger D.J., Arnot D.E., Tam J.P., Nussenzweig V., Enea V.;  
RT "Circumsporozoite protein of Plasmodium berghei: gene cloning and  
RT identification of the immunodominant epitopes.";  
RL Mol. Cell. Biol. 6:3965-3972(1986).  
CC -1- FUNCTION: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT  
CC SURFACE ANTIGEN ON THE SPOOROZOITE (THE INFECTIVE STAGE OF THE  
CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE

```
CC VERTEBRATE HOST).  
CC - MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR  
CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES  
CC WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.  
CC -----  
CC - SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; M14135; AAA29577.1; .  
DR PIR; A25083; OZKOMB.  
DR InterPro; IPR003067; Crmsprztoite.  
DR InterPro; IPR000884; TSPL.  
DR Pfam; PF00090; tsp_1; 1.  
DR PRINTS; PR01303; CRMSPRZOITE.  
DR SMART; SMO0209; TSPI; 1.  
DR PROSITE; PS50092; TSPI; 1.  
KW Malaria; Sporozoite; Repeat; Signal.  
FT SIGNAL 1 23 PROBABLE.  
FT CHAIN 24 339 CIRCUMSPOROZITE PROTEIN.  
FT DOMAIN 93 196 13 X 8 AA TANDEM REPEATS.  
FT DOMAIN 206 238 16 X 2 AA TANDEM REPEATS OF P-Q.  
SQ SEQUENCE 339 AA; 37138 MW; E8068A6D1D9551B CRC64;  
  
QY Query Match 46.1%; Score 58.5; DB 1; Length 339;  
Best Local Similarity 52.2%; Pred. No. 0.71;  
Matches 12; Conservative 5; Mismatches 5; Indels 1; Gaps 1;  
  
Db 1 KPNDELVDNDIEKKICKMKCS 23  
296 KRAEDLTLE-DIDTEICKMDKCS 317  
  
RESULT 12  
CSP_PLABA STANDARD; PRT; 347 AA.  
ID CSP_PLABA  
AC P23093;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE CIRCUMSPOROZITE PROTEIN PRECURSOR (CS).  
OS Plasmodium berghei (strain Anka) .  
OX Eukaryota; Alveolata; Apicomplexa; Hemosporidia; Plasmodium.  
NCBI_TaxID=5823;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90221834; PubMed=2183186;  
RA Lockyer M.J., Davies C.S., Subdier A., Sindgen R.E.;  
RL "Nucleotide sequence of the Plasmodium berghei circumsporozoite  
protein gene from the ANKA clone 2.34f.";  
RT Nucleic Acids Res. 18:376-376(1990).  
RT FUNCTION: THE CIRCUMSPOROZITE PROTEIN IS THE IMMUNODOMINANT  
RT -I- SURFACE ANTIGEN ON THE SPOROZYTE (THE INFECTIVE STAGE OF THE  
CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE  
CC VERTEBRATE HOST).  
CC -I- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR  
CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES  
CC WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.  
CC -----  
CC -I- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).
```

```

DR EMBL: X17606; CAA35608.1; -.
DR PIR: S07873; OZDOBK.
DR InterPro: IPR003067; Circmsporozite.
DR InterPro: IPR000884; TSP1.
DR Pfam: PF00090; tsp.1; 1.
DR PRINTS: PR01303; CircmSPROITE.
DR SMART: SM00209; TSP1.1.
DR PROSITE: PS50092; TSP1.
KW Malaria; Sporozite; Repeat; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 347 CIRCUMSPOROZITE PROTEIN.
FT DOMAIN 93 204 13 X 8 AA REPEATS.
FT DOMAIN 214 247 17 X 2 AA REPEATS OF P-Q.
SQ SEQUENCE 347 AA; 37776 MW; 0EC240EE35681AF8 CRC64;

Query Match 46.1%; Score 58.5; DB 1; Length 347;
Best Local Similarity 52.2%; Pred. No. 0.73;
Matches 12; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

OY 1 RKPDLDYENDIEKKICKMKRCS 23
|::| |::| |::| |::|
DB 304 KKAEDTLE-DIDETICKMKRCS 325

RESULT 13
CSP_PLAYO STANDARD; PRT; 367 AA.
AC P06914;
DT 01-JAN-1988 (Rel. 06, Created)
DT 20-JAN-1988 (Rel. 06, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CIRCUMSPOROZITE PROTEIN PRECURSOR (CS).
OS Plasmodium berghei yoelli.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5862;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87137555; PubMed=3102479;
RA Lal A.A., de la Cruz V.F., Welsh J.A., Charoenvit V., Maloy W.L.,
RA McCutchan T.F.;
RT "Structure of the gene encoding the circumsporozoite protein of
RT Plasmodium yoelli. A rodent model for examining antimalarial
RT sporozoite vaccines."
RL J. Biol. Chem. 262:2937-2940(1987).
RN [2]
RP SEQUENCE OF 1-140 AND 260-367 FROM N.A.
RX MEDLINE=88232798; PubMed=3287156;
RA de la Cruz V.F., Lal A.A., McCutchan T.F.;
RT "Variation among circumsporozoite protein genes from rodent
RT malaria."
RL Mol. Biochem. Parasitol. 28:31-38(1988).
RN [3]
RP FUNCTION: THE CIRCUMSPOROZITE PROTEIN IS THE IMMUNODOMINANT
RP SURFACE ANTIGEN ON THE SPOROZITE (THE INFECTIVE STAGE OF THE
RP MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE
RP VERTEBRATE HOST).
RN [4]
RP MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR
RP ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES
RP WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.
RN [5]
RP SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation on
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC -----
DR EMBL: J02695; AAA29558.1; -.
DR EMBL: M18821; AAA29559.1; -.
DR EMBL: M22698; AAA29560.1; -.

```



---



ID Q90P8 PRELIMINARY; PRT; 80 AA.  
AC Q90P8;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE CIRCUMSPOROZOITE PROTEIN (FRAGMENT).  
GN CS.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OC NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=M4.  
RA de Stricker K., Vuust J., Jepsen S., Oeuvray C., Theisen M.;  
RT "Sequence variation in the non-repeat region of the Plasmodium  
RT falciparum glutamate rich protein (GLURP) from Brazil, Senegalese, and  
RT Burmese field isolates and from laboratory strains."  
RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AJ269957; CAB64237.1; -.  
DR InterPro: IPR003067; Circsprzote.  
DR Pfam: PF00090; tsp\_1.1  
DR PRINTS: PR01303; CRCMSPRZOITE.  
DR SMART: SM00209; TSP1.1.  
FT NON\_TER 1 1  
FT NON\_TER 80 80  
SQ SEQUENCE 80 AA; 9073 MW; ABF04B8FB142B1E CRC64;

Query Match 100.0%; Score 127; DB 5; Length 80;  
Best Local Similarity 100.0%; Pred. No. 1.1e-09;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPDELVDYNDIEKKICKMEKCS 23  
Db 50 KPDELVDYNDIEKKICKMEKCS 72

RESULT 3  
ID Q90P7 PRELIMINARY; PRT; 80 AA.  
AC Q90P7;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE CIRCUMSPOROZOITE PROTEIN (FRAGMENT).  
GN CS.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OC NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=D4230;  
RA de Stricker K., Vuust J., Jepsen S., Oeuvray C., Theisen M.;  
RT "Sequence variation in the non-repeat region of the Plasmodium  
RT falciparum glutamate rich protein (GLURP) from Brazil, Senegalese, and  
RT Burmese field isolates and from laboratory strains."  
RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AJ269961; CAB64180.1; -.  
DR InterPro: IPR000884; TSP1.  
DR InterPro: IPR003067; Circsprzote.  
DR Pfam: PF00090; tsp\_1.1.  
DR PRINTS: PR01303; CRCMSPRZOITE.  
DR SMART: SM00209; TSP1.1.  
FT NON\_TER 1 1  
FT NON\_TER 80 80  
SQ SEQUENCE 80 AA; 9047 MW; BA769C90DB031C3E CRC64;

Query Match 100.0%; Score 127; DB 5; Length 80;  
Best Local Similarity 100.0%; Pred. No. 1.1e-09;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPDELVDYNDIEKKICKMEKCS 23  
Db 50 KPDELVDYNDIEKKICKMEKCS 72

RESULT 4  
ID Q90P3 PRELIMINARY; PRT; 80 AA.  
AC Q90P3;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE CIRCUMSPOROZOITE PROTEIN (FRAGMENT).  
GN CS.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OC NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=D4405;  
RA de Stricker K., Vuust J., Jepsen S., Oeuvray C., Theisen M.;  
RT "Sequence variation in the non-repeat region of the Plasmodium  
RT falciparum glutamate rich protein (GLURP) from Brazil, Senegalese, and  
RT Burmese field isolates and from laboratory strains."  
RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AJ269969; CAB64188.1; -.  
DR InterPro: IPR000884; TSP1.  
DR InterPro: IPR003067; Circmsprzote.  
DR Pfam: PF00090; tsp\_1.1.  
DR PRINTS: PR01303; CRCMSPRZOITE.  
DR SMART: SM00209; TSP1.1.  
FT NON\_TER 1 1  
FT NON\_TER 80 80  
SQ SEQUENCE 80 AA; 9033 MW; BA71EBEDDB03193B CRC64;

Query Match 100.0%; Score 127; DB 5; Length 80;  
Best Local Similarity 100.0%; Pred. No. 1.1e-09;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPDELVDYNDIEKKICKMEKCS 23  
Db 50 KPDELVDYNDIEKKICKMEKCS 72

RESULT 5  
ID Q90P1 PRELIMINARY; PRT; 80 AA.  
AC Q90P1;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE CIRCUMSPOROZOITE PROTEIN (FRAGMENT).  
GN CS.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OC NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B1896;  
RA de Stricker K., Vuust J., Jepsen S., Oeuvray C., Theisen M.;  
RT "Sequence variation in the non-repeat region of the Plasmodium  
RT falciparum glutamate rich protein (GLURP) from Brazil, Senegalese, and  
RT Burmese field isolates and from laboratory strains."  
RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AJ269978; CAB64197.1; -.  
DR InterPro: IPR003067; Circmsprzote.  
DR InterPro: IPR000884; TSP1.  
DR Pfam: PF00090; tsp\_1.1  
DR PRINTS: PR01303; CRCMSPRZOITE.  
DR SMART: SM00209; TSP1.1.  
FT NON\_TER 1 1  
FT NON\_TER 80 80

50	SEQUENCE	80	AA:	9102	MM:	BFC6C970CE0FPA3E	CRC64:	
	Query Match			100.0%:	Score 127:	DB 5:	Length 80:	
	Best Local Similarity			100.0%:	Pred. No. 1,1e-09:			
	Matches	23:	Conservative	0:	Mismatches	0:	Indels	
OY	1	KPKDELDYENDIEKKICKMKCS	23					
DB	50	KPKDELDYENDIEKKICKMKCS	72					
RESULT	6							
O9TVN9		PRELIMINARY:		PRT:	80	AA:		
ID	O9TVN9							
AC	O9TVN9:							
DT	01-MAY-2000 (TREMBLrel. 13, Created)							
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)							
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)							
DE	CIRCUMSPOROZOITE PROTEIN (FRAGMENT).							
DE	CS.							
GN	Plasmodium falciparum.							
OS	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.							
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.							
OX	NCBI_TaxID=5833;							
RN	[1]							
RP	SEQUENCE FROM N.A.							
RC	STRAIN-M0, M6, M7, AND M1;							
RC	de Stricker K., Vuust J., Jepsen S., Oeuvray C., Theisen M.;							
RT	"Sequence variation in the non-repeat region of the Plasmodium							
RT	falciparum glutamate rich protein (GLURP) from Brazil, Senegalese, and							
RT	Burmese field isolates and from laboratory strains."							
RL	Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.							
DR	EMBL; AJ269595; CAB64242.1; -							
DR	EMBL; AJ269598; CAB64238.1; -							
DR	EMBL; AJ269599; CAB64239.1; -							
DR	EMBL; AJ269596; CAB64241.1; -							
DR	InterPro: IPR000884; TSP1.							
DR	InterPro: IPR003067; Circspprzoite.							
DR	Pfam: PF00090; TSP_1; 1.							
DR	PRINTS: PRO1303; CRCMSPRZOITE.							
DR	SMART: SMO0209; TSP1. 1.							
FT	NON_TER	1						
FT	NON_TER	80						
SO	SEQUENCE	80	AA:	9074	MM:	A8F40C90DB1C033E	CRC64:	
	Query Match			100.0%:	Score 127:	DB 5:	Length 80:	
	Best Local Similarity			100.0%:	Pred. No. 1,1e-09:			
	Matches	23:	Conservative	0:	Mismatches	0:	Indels	
OY	1	KPKDELDYENDIEKKICKMKCS	23					
DB	50	KPKDELDYENDIEKKICKMKCS	72					
RESULT	7							
O9TVN9		PRELIMINARY:		PRT:	80	AA:		
ID	O9TVN9							
AC	O9TVN9:							
DT	01-MAY-2000 (TREMBLrel. 13, Created)							
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)							
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)							
DE	CIRCUMSPOROZOITE PROTEIN (FRAGMENT).							
DE	CS.							
GN	Plasmodium falciparum.							
OS	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.							
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.							
OX	NCBI_TaxID=5833;							
RN	[1]							
RP	SEQUENCE FROM N.A.							
RC	STRAIN-B1893, B1804, B1853, B1870, B1872, B1882, B1887,							
RA	de Stricker K., Vuust J., Jepsen S., Oeuvray C., Theisen M.;							
RT	"Sequence variation in the non-repeat region of the Plasmodium							
RT	falciparum glutamate rich protein (GLURP) from Brazil, Senegalese, and							

```

RT   Submitted (SFP-1999) to the EMBL/GenBank/DBJ databases.
RL   EMBL: AJ269977; CAB64196.1; -
DR   EMBL: AJ269971; CAB64190.1; -
DR   EMBL: AJ269972; CAB64191.1; -
DR   EMBL: AJ269973; CAB64192.1; -
DR   EMBL: AJ269974; CAB64193.1; -
DR   EMBL: AJ269975; CAB64194.1; -
DR   EMBL: AJ269976; CAB64195.1; -
DR   InterPro: IPR000884; TSP1.
DR   InterPro: IPR003067; Crcmsprzotite.
DR   Pfam: PF00090; TSP_1; 1.
DR   PRINTS: PR01303; Crcmsprzotite.
DR   SMART: SM00209; TSP1; 1.
FT   NON_TER 1
FT   NON_TER 80
SQ   SEQUENCE 80 AA; 9074 MW; ADF10C8FC41C1924 CRC64;

Query Match 100.0%; Score 127; DB 5; Length 80;
Best Local Similarity 100.0%; Pred. No. 1,1e-09;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPKDELDYENDIEKKICKMEKCS 23
DB 50 KPKDELDYENDIEKKICKMEKCS 72

RESULT 8
O9TVM7 PRELIMINARY; PRT: 115 AA.
O9TVM7
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CIRCUMSPOROZOITE PROTEIN (FRAGMENT).
GN Csp.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=815, 947, AND 808;
RX MEDLINE=95077069; PubMed=7985759;
RA Jongwattien S., Tanabe K., Hughes M.K., Kanbara H., Hughes A.L.;
RT "Allelic variation in the circumsporozoite protein of Plasmodium
   falciparum from Thai field isolates.";
RL Am. J. Trop. Med. Hyg. 51:659-668(1994).
DR EMBL: M83153; AAA29566.1; -
DR EMBL: M83171; AAA29549.1; -
DR EMBL: M83151; AAA29564.1; -
DR InterPro: IPR000884; TSP1.
DR InterPro: IPR003067; Crcmsprzotite.
DR Pfam: PF00090; TSP_1; 1.
DR PRINTS: PR01303; Crcmsprzotite.
DR SMART: SM00209; TSP1; 1.
FT   NON_TER 1
FT   NON_TER 1
SQ   SEQUENCE 115 AA; 12925 MW; 697DF63EEBBDAA90 CRC64;

Query Match 100.0%; Score 127; DB 5; Length 115;
Best Local Similarity 100.0%; Pred. No. 1,1e-09;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPKDELDYENDIEKKICKMEKCS 23
DB 71 KPKDELDYENDIEKKICKMEKCS 93

RESULT 9
O27425 PRELIMINARY; PRT: 424 AA.
O27425
AC 027425;

```

```
DR 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CIRCUMSPOROZOITE PROTEIN.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=844;
RX MEDLINE=84250215; PubMed=6204383;
RA Dame J.B., Williams J.L., McCutchan T.F., Wirtz R.A.,
RA Hockmeyer W.T., Maloy W.L., Haynes J.D., Schneider I., Roberts D.,
RA Sanders G.S., Reddy P.E., Diggs C.L., Miller L.H.;
RT "Structure of the gene encoding the immunodominant surface antigen on
RT the sporozoite of the human malaria parasite Plasmodium falciparum.";
RL Science 225:593-599(1984).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=844;
RA la Cruz V.F.;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN=844;
RX Jongwutlives S., Tanabe K., Kanbara H.;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; M83169; AAA29547.1; -
DR EMBL; M83169; AAA29562.1; -
DR InterPro: IPR003067; Crcmsprzoite.
DR InterPro: IPR003067; Crcmsprzoite.
DR Pfam: PF00090; tsp_1; 1
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
DR PRINTS: PR01303; CIRCSPRZOITE.
DR SMART: SM00209; TSP1.1.
SQ SEQUENCE 424 AA; 45592 MW; F20CEB60636DB98E CRC64;

Query Match 100.0%; Score 127; DB 5; Length 424;
Best Local Similarity 100.0%; Pred. No. 5e-09;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPKDELDYENDIEKICKMEKCS 23
DB 380 KPKDELDYENDIEKICKMEKCS 402

RESULT 10
O99256 PRELIMINARY; PRT; 424 AA.
AC O99256;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CIRCUMSPOROZOITE PROTEIN (CS) PRECURSOR, VARIANT 2 (FRAGMENT).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=844;
RX MEDLINE=91270295; PubMed=2052038;
RA Lockyer M.J.;
RT "Clonal variation in the Plasmodium falciparum circumsporozoite
RT protein gene.";
RL Mol. Biochem. Parasitol. 45:179-182(1991).
CC -1- FUNCTION: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT
CC SURFACE ANTIGEN ON THE SPOROZOITE (THE INFECTIVE STAGE OF THE
CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE
CC VERTEBRATE HOST).
CC -1- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR
CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES
CC WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.
DR EMBL; M57499; AAA63422.1; -
```

```
DR InterPro: IPR000884; TSP1.
DR InterPro: IPR003067; Crcmsprzoite.
DR Pfam: PF00090; tsp_1; 1
DR PRINTS: PR01303; CIRCSPRZOITE.
DR SMART: SM00209; TSP1.1.
KW Malaria; Repeat; Sporozoite; Signal.
FT SIGNAL 1 16 BY SIMILARITY.
FT CHAIN 17 >424 CIRCUMSPOROZOITE PROTEIN.
FT DOMAIN 123 302 4-RESIDUE TANDEM REPEATS.
FT NON-TER 424 424
SQ SEQUENCE 424 AA; 45609 MW; BFD0F906C664B8F CRC64;

Query Match 100.0%; Score 127; DB 5; Length 424;
Best Local Similarity 100.0%; Pred. No. 5e-09;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPKDELDYENDIEKICKMEKCS 23
DB 380 KPKDELDYENDIEKICKMEKCS 402

RESULT 11
O27246 PRELIMINARY; PRT; 432 AA.
AC O27246;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CIRCUMSPOROZOITE PROTEIN.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=946;
RX MEDLINE=84250215; PubMed=6204383;
RA Dame J.B., Williams J.L., McCutchan T.F., Weber J.L., Wirtz R.A.,
RA Hockmeyer W.T., Maloy W.L., Haynes J.D., Schneider I., Roberts D.,
RA Sanders G.S., Reddy P.E., Diggs C.L., Miller L.H.;
RT "Structure of the gene encoding the immunodominant surface antigen on
RT the sporozoite of the human malaria parasite Plasmodium falciparum.";
RL Science 225:593-599(1984).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=946;
RA la Cruz V.F.;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN=946;
RX Jongwutlives S., Tanabe K., Kanbara H.;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; M83155; AAA29568.1; -
DR EMBL; M83170; AAA29548.1; -
DR EMBL; M83152; AAA29565.1; -
DR EMBL; M83158; AAA29571.1; -
DR EMBL; M83166; AAA29544.1; -
DR EMBL; M83168; AAA29546.1; -
DR InterPro: IPR000884; TSP1.
DR InterPro: IPR003067; Crcmsprzoite.
DR Pfam: PF00090; tsp_1; 1
DR PRINTS: PR01303; CIRCSPRZOITE.
DR SMART: SM00209; TSP1.1.
SQ SEQUENCE 432 AA; 46414 MW; 8787E6005578873A CRC64;

Query Match 100.0%; Score 127; DB 5; Length 432;
Best Local Similarity 100.0%; Pred. No. 5.1e-09;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPKDELDYENDIEKICKMEKCS 23
```



Db 388 KPKDELVDYNDIEKKICKMEKCS 410

## RESULT 12

ID Q25827 PRELIMINARY; PRT: 432 AA.

AC Q25827; 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE CIRCUMSPOROZOITE PROTEIN.

GN CSP.

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI\_TaxID=5633;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=838;

RX MEDLINE=84250215; PubMed=6204383;

RA Dame J.B., Williams J.L., McCutchan T.F., Weber J.L., Wirtz R.A.,

Hockmeyer W.T., Maloy W.L., Haynes J.D., Schneider I., Roberts D.,

Sanders G.S., Reddy P.E., Diggs C.L., Miller L.H.;

RT "Structure of the gene encoding the immunodominant surface antigen on

the sporozoite of the human malaria parasite Plasmodium falciparum.";

RL Science 225:593-599(1984).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=838;

RA la Cruz V.F.;

RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=838;

RA Jongwutlives S., Tanabe K., Kanbara H.;

RL Mol. Biochem. Parasitol. 0:0-0(0).

DR EMBL: M83165; AAA29543.1;

DR InterPro: IPR000884; TSP1.

DR InterPro: IPR003067; Circumsporzoite.

DR Pfam: PF00090; TSP1.1;

DR PRINTS: PR01303; CRCMSPRZOITE.

DR SMART: SM00209; TSP1.1;

SO SEQUENCE 432 AA; 46385 MW; 2CB8D9A68B11945F CRC64;

OY 1 KPKDELVDYNDIEKKICKMEKCS 23

Db 388 KPKDELVDYNDIEKKICKMEKCS 410

Query Match 100.0%; Score 127; DB 5; Length 432;

Best Local Similarity 100.0%; Pred. No. 5.1e-09;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## RESULT 13

ID Q9U001 PRELIMINARY; PRT: 80 AA.

AC Q9U001; 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE CIRCUMSPOROZOITE PROTEIN (FRAGMENT).

GN CS.

OS Plasmodium falciparum (isolate K1 / Thailand).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI\_TaxID=5633;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=K1;

RA de Stricker K., Vuust J., Jepsen S., Oeuvray C., Theisen M.;

RT "Sequence variation in the non-repeat region of the Plasmodium

falciparum glutamate rich protein (GURP) from Brazil, Senegalese, and

Burmese field isolates and from laboratory strains.";

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AJ269946; CAB64171.1;

DR InterPro: IPR000884; TSP1.

DR InterPro: IPR003067; Circumsporzoite.

DR Pfam: PF00090; TSP1.1;

DR PRINTS: PR01303; CRCMSPRZOITE.

DR PROSITE: PS50092; TSP1.1.

DR SMART: SM00209; TSP1.1.

FT NON\_TER 1

FT NON\_TER 80

SO SEQUENCE 80 AA; 9137 MW; 4204EBC6F0B1434 CRC64;

Query Match 97.6%; Score 124; DB 5; Length 80;

Best Local Similarity 95.7%; Pred. No. 2.7e-09;

Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPKDELVDYNDIEKKICKMEKCS 23

Db 50 KPKDELVDYNDIEKKICKMEKCS 72

## RESULT 14

ID Q9U000 PRELIMINARY; PRT: 80 AA.

AC Q9U000; 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE CIRCUMSPOROZOITE PROTEIN (FRAGMENT).

GN CS.

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI\_TaxID=5633;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=D28;

RA de Stricker K., Vuust J., Jepsen S., Oeuvray C., Theisen M.;

RT "Sequence variation in the non-repeat region of the Plasmodium

falciparum glutamate rich protein (GURP) from Brazil, Senegalese, and

Burmese field isolates and from laboratory strains.";

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AJ269948; CAB64173.1;

DR InterPro: IPR000884; TSP1.

DR InterPro: IPR003067; Circumsporzoite.

DR Pfam: PF00090; TSP1.1;

DR PRINTS: PR01303; CRCMSPRZOITE.

DR SMART: SM00209; TSP1.1.

FT NON\_TER 1

FT NON\_TER 80

SO SEQUENCE 80 AA; 9046 MW; BA7689D18F031C3E CRC64;

Query Match 97.6%; Score 124; DB 5; Length 80;

Best Local Similarity 95.7%; Pred. No. 2.7e-09;

Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPKDELVDYNDIEKKICKMEKCS 23

Db 50 KPKDELVDYNDIEKKICKMEKCS 72

## RESULT 15

ID Q9U0P9 PRELIMINARY; PRT: 80 AA.

AC Q9U0P9; 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE CIRCUMSPOROZOITE PROTEIN (FRAGMENT).

GN CS.

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1;  
RA de Stricker K., Vuust J., Jepsen S., Ouevray C., Theisen M.;  
RT "Sequence variation in the non-repeat region of the plasmodium  
RT falciparum glutamate rich protein (GLURP) from Brazil, Senegalese, and  
RT Burmese field isolates and from laboratory strains."  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ269951; CAB64176.1; -  
DR InterPro: IPR003067; Crmsprzoite.  
DR Pfam: PF00090; tsp\_1; 1.  
DR PRINTS: PR01303; CRGMSPRZOITE.  
DR SMART: SM00209; TSP1; 1.  
FT NON\_TER 1  
FT NON\_TER 80  
SQ SEQUENCE 80 AA: 9032 MW: ADED6F0E266AD98E CRC64;

Query Match 97.6%; Score 124; DB 5; Length 80;  
Best Local Similarity 95.7%; Pred. No. 2.7e-09;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KPREDLDYENDIEKKICKMEKCS 23  
|||||  
DB 50 KPRDQLDYENDIEKKICKMEKCS 72

Search completed: January 29, 2002, 11:12:10  
Job time: 766 sec



XX Novel recombinant protein as vaccine for treating malarial infection  
 PT comprises antigenic peptides obtained from different stages of  
 PT Plasmodium falciparum life cycle  
 PS Claim 2; Page 16; 52pp; English.  
 XX  
 CC The present sequence is the antigenic epitope P594, derived from  
 CC circumsporozoite protein (CSP) of the sporozoite stage of Plasmodium  
 CC falciparum. It is used in the construction of recombinant protein.  
 CC CDC/NITMALVAC-1, which is a multivalent, multistage malarial vaccine.  
 CC The recombinant protein comprises, melittin signal peptide, (His)<sub>6</sub> tag,  
 CC T-cell epitope from tetanus toxoid and 21 antigenic epitopes from  
 CC circumsporozoite protein (CSP), sporozoite surface protein-2 (SSP-2),  
 CC liver stage antigen-1 (LSA-1), merozoite surface protein-1 (MSP-1),  
 CC MSP-2, apical membrane antigen-1 (AMA-1), erythrocyte binding antigen-175  
 CC (EBA-175), rhoptry associated protein-1 (RAP-1) and gamete specific  
 CC antigen, Pf27. These epitopes were obtained at different stages of the  
 CC life cycle of P. falciparum. CDC/NITMALVAC-1 vaccine has antiparasitic  
 CC activity and can be used for treatment and prevention of malarial  
 CC infections. Anti-CDC/NITMALVAC-1 antibodies can be used for detecting  
 CC P. falciparum in biological samples.  
 CC  
 XX  
 SQ Sequence 21 AA:

Query Match 100.0%; Score 109; DB 21; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 4.4e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DIEKKICKMEKCSVFNVNVS 21  
 DB 1 diekkickmekcsvfnyvns 21  
 ||||||||||||||||  
 ||||||||||||||||

RESULT 2  
 AAR13175  
 ID AAR13175 standard; Protein; 309 AA.  
 XX  
 AC AAR13175;  
 XX  
 DT 29-AUG-1991 (first entry)  
 XX  
 DE NS1\_81-RLfdelta9.  
 XX  
 KW Immunogenic determinant; circumsporozoite; CS; vaccine; malaria;  
 KW hybrid; influenza virus; non-structural protein 1; fusion.  
 XX  
 OS Plasmodium falciparum.  
 OS Influenza virus (A/PK/8/34/).

XX  
 FH Key Location/Qualifiers  
 FT 1..81  
 FT /label= N-terminal of NS1  
 FT /note= "Influenza virus nonstructural protein 1"  
 FT 82..87  
 FT /label= synthetic linker  
 FT 88  
 FT /label= artifact  
 FT /note= "see comments"  
 FT 89..193  
 FT /label= AAS 19-123 of CS protein  
 FT /note= "Region I contg. flanking region less  
 FT 194..309  
 FT /label= AAS 297-412 of CS protein  
 FT /note= "Region II flanking region minus 9 N-term-  
 FT inal AAS"  
 FT  
 PN EP432965-A.  
 XX  
 XX 19-JUN-1991.  
 PD  
 XX

PF 06-DEC-1990; 90EP-0313257.  
 XX  
 PR 08-DEC-1989; 89US-0447746.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM.  
 PA (USA ) US SEC OF THE ARMY.  
 PA (BIOM-) BIOMEDICAL RES INST.  
 XX

PI Gross MS, Gordon DM, Hollingdale MR;  
 DR WPI; 1991-179771/25.  
 XX

PT Polypeptide comprising immunogenic determinants from P falciparum  
 PT - for vaccine against malaria infection in humans.  
 PS Example 1; Page 7; 18pp; English.

XX The polypeptide is prepd. by genetic engineering of genes encoding  
 CC the P. falciparum circumsporozoite (CS) protein [Dane et al.,  
 CC Science 225 : 593 (1984)], and the influenza virus non-structural  
 CC protein 1 (NS1), [Baez et al., Nucleic Acids Research, 8 : 5845  
 CC (1980)]. The DNA encoding the 1st 81 AAs of the N-terminal of NS1  
 CC (NS1\_81) is linked via a synthetic sequence to DNA encoding Region  
 CC I contg. flanking regionless the 18 AA signal region, which in  
 CC turn is fused to DNA encoding Region II-contg. flanking region  
 CC less the first nine N-terminal AAs. This CS fusion is designated  
 CC Rfdelta9. The pro residue separating the Asp (at the C-terminal  
 CC of the linker) from Rfdelta9 is an artifact of a filled in BamHI  
 CC site. The peptide can be used in a vaccine for protection against  
 CC malaria.  
 CC See also AAR12306-RI2311 and AAR13176-RI3179.  
 CC  
 XX  
 SQ Sequence 309 AA:

Query Match 100.0%; Score 109; DB 12; Length 309;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-09;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DIEKKICKMEKCSVFNVNVS 21  
 DB 275 diekkickmekcsvfnyvns 295  
 ||||||||||||||||  
 ||||||||||||||||

RESULT 3  
 AAR07945  
 ID AAR07945 standard; Protein; 319 AA.  
 XX  
 AC AAR07945;  
 XX  
 DT 22-FEB-1991 (first entry)  
 XX  
 DE NS181RLFAuth plasmid product.  
 XX  
 KW Malaria; vaccine.  
 XX  
 OS Plasmodium falciparum.

XX  
 FH Key Location/Qualifiers  
 FT 1..81  
 FT /label= NS181 protein fragment  
 FT /note= "from plasmid pmG-1"  
 FT 89..193  
 FT /label= Fragment of circumsporozoite protein  
 FT 204..319  
 FT /label= Fragment of circumsporozoite protein  
 FT  
 PN EP398540-A.  
 XX  
 XX 22-NOV-1990.  
 PD  
 XX 01-MAY-1990; 90EP-0304720.  
 PF  
 XX

```

PR 03-MAY-1989; 89US-0346833.
XX
XX PA (SMIK ) SMITHKLINE BEECHAM.
XX
XX PI Gross MS, Young JF;
XX
XX DR WPI: 1990-350299/47.
XX N-PSDB: AAQ06580.
XX
XX PT New polypeptide used in malaria vaccine - comprises immunogenic
XX determinants from 1st and 2nd flanking regions of Plasmodium
XX surface protein and intermediate repeat domain
XX
XX PS Example 2; Page 11-12; 24pp: English.
XX
XX CC The product is useful in preparation of vaccines for treatment and
XX prophylaxis of plasmodium sporozite infection. It may be easily
XX produced in large pure quantities from a transformed E.coli
XX expression system.
XX
XX CC
XX
XX SO Sequence 319 AA;

Query Match 100.0%; Score 109; DB 11; Length 319;
Best Local Similarity 100.0%; Pred. No. 7.4e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 1 DIEKICKMEKCSYFNVVNS 21
   ||||||||||||||||
Db 285 diekickmekcsyfnvns 305

RESULT 4
AAR13176
ID AAR13176 standard; Protein; 319 AA.
XX
XX AC AAR13176;
XX
XX DT 29-AUG-1991 (first entry)
XX
XX DE NS1_81-RLFAuth.
XX
XX KW Immunogenic determinant; circumsporozoite; CS; vaccine; malaria;
XX hybrid; Influenza virus; non-structural protein 1; fusion.
XX
XX OS Plasmodium falciparum.
XX OS Influenza virus (A/PR/8/34).
XX
XX FH Location/Qualifiers
XX Key 1..81
XX Region /label= N-terminal of NS1
XX FT /note= "Influenza virus nonstructural protein 1"
XX FT Peptide 82..87
XX FT /label= synthetic linker
XX FT Region 88
XX FT /label= artifact
XX FT /note= "see comments"
XX FT Region 89..193
XX FT /label= AAs 19-123 of CS protein
XX FT /note= "Region I contg. flanking region less
XX signal sequence"
XX FT Region 194
XX FT /label= artifact
XX FT /note= "see comments"
XX FT Region 195..319
XX FT /label= AAs 288-412 of CS protein
XX FT /note= "Region II flanking region"
XX
XX PN EPA32965-A.
XX
XX PD 19-JUN-1991.
XX
XX 06-DEC-1990; 90EP-0313257.

```

[illegible]

```

FT      /note="Region I contg. flanking region less
FT      signal sequence"
FT      Region
FT      194..201
FT      /label="immunodominant repeat region"
FT      /note="two tetrapeptide repeat units"
FT      Region
FT      202
FT      /label="artifact"
FT      /note="see comments"
FT      203..327
FT      /label="AAs 288-412 of CS protein"
FT      /note="Region II flanking region"
FT
FT
FT      EP432965-A.
FT      19-JUN-1991.
FT
FT      06-DEC-1990; 90EP-0313257.
FT
FT      08-DEC-1989; 89US-0447746.
FT
FT      (SMIK ) SMITHKLINE BEECHAM.
FT      (USSA ) US SEC OF THE ARMY.
FT      (BIOM-) BIOMEDICAL RES INST.
FT
FT      Gross MS, Gordon DM, Hollingdale MR;
FT      WPI; 1991-179771/25.
FT
FT      Polypeptide comprising immunogenic determinants from P falciparum
FT      - for vaccine against malaria infection in humans.
FT
FT      Example 3; Page 10; 18pp; English.
FT
FT      The polypeptide is prep'd. by genetic engineering of genes encoding
FT      the P. falciparum circumsporozoite (CS) protein [Dame et al.,
FT      Science 225 : 593 (1984)] and the influenza virus non-structural
FT      protein 1 (NS1), [Baez et al., Nucleic Acids Research, 8 : 5845
FT      (1980)]. The DNA encoding the 1st 81 AAs of the N-terminal of NS1
FT      (NS1.81) is linked via a synthetic sequence to DNA encoding Region
FT      I contg. flanking region less the 18 AA signal region. This is
FT      linked to a synthetic sequence encoding two repeat units from the
FT      immunodominant region, which in turn is fused to DNA encoding
FT      the Region II contg. flanking region. The Pro residue separating the
FT      Asp (at the C-terminal of the linker) from the Region I contg. CS
FT      flanking region is an artifact of a filled-in BamHI site; the Gly
FT      separating the repeat units and the Region II contg. CS flanking
FT      region is an artifact of a synthetic FokI/ThiIII I linker. The
FT      peptide can be used in a vaccine for protection against malaria.
FT      See also AAR12306-R12311 and AAR13175-R13179.
FT
FT      Sequence 327 AA;
FT
FT      Query Match          100.0%; Score 109; DB 12; Length 327;
FT      Best Local Similarity 100.0%; Pred. No. 7.5e-09;
FT      Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
FT
FT      QY      1 DIEKRIKMEKCSVFNVNS 21
FT      ||||||||||||||||
FT      DB      293 diekrikmekcsvfnvns 313
FT
FT      RESULT 6
FT      AAR13178
FT      ID      AAR13178 standard; Protein: 335 AA.
FT      XX
FT      AC      AAR13178;
FT      XX
FT      DT      29-AUG-1991 (first entry)
FT      XX
FT      DE      NS1_81(NANP)4RLFAuth.
FT      XX
FT      KW      Immunogenic determinant; circumsporozoite; CS; vaccine; malaria;

```

```

KW      hybrid; influenza virus; non-structural protein 1; fusion.
XX
XX      OS      Plasmodium falciparum.
OS      Influenza virus (A/PR/8/34).
XX
XX      Key      Location/Qualifiers
FH      Region      1..81
FT      /label="N-terminal of NS1"
FT      /note="Influenza virus nonstructural protein 1"
FT      82..97
FT      /label="immunodominant repeat region"
FT      /note="four tetrapeptide repeat units"
FT      98..103
FT      /label="synthetic linker"
FT      104
FT      /label="artifact"
FT      /note="see comments"
FT      105..209
FT      /label="AAs 19-123 of CS protein"
FT      /note="Region I contg. flanking region less
FT      signal sequence"
FT      210
FT      /label="artifact"
FT      /note="see comments"
FT      211..335
FT      /label="AAs 288-412 of CS protein"
FT      /note="Region II flanking region"
FT
FT      EP432965-A.
FT      19-JUN-1991.
FT
FT      06-DEC-1990; 90EP-0313257.
FT
FT      08-DEC-1989; 89US-0447746.
FT
FT      (SMIK ) SMITHKLINE BEECHAM.
FT      (USSA ) US SEC OF THE ARMY.
FT      (BIOM-) BIOMEDICAL RES INST.
FT
FT      Gross MS, Gordon DM, Hollingdale MR;
FT      WPI; 1991-179771/25.
FT
FT      Polypeptide comprising immunogenic determinants from P falciparum
FT      - for vaccine against malaria infection in humans.
FT
FT      Example 4; Page 11; 18pp; English.
FT
FT      The polypeptide is prep'd. by genetic engineering of genes encoding
FT      the P. falciparum circumsporozoite (CS) protein [Dame et al.,
FT      Science 225 : 593 (1984)] and the influenza virus non-structural
FT      protein 1 (NS1), [Baez et al., Nucleic Acids Research, 8 : 5845
FT      (1980)]. The DNA encoding the 1st 81 AAs of the N-terminal of NS1
FT      (NS1.81) is linked to a synthetic sequence encoding four repeat
FT      units from the immunodominant region, which in turn is linked via
FT      a synthetic sequence to DNA encoding Region I contg. flanking
FT      region less the 18 AA signal region. This is linked to DNA
FT      encoding Region II contg. flanking region. The Pro residue sep-
FT      arating the Asp (at the C-terminal of the linker) from the Region
FT      I contg. CS flanking region is an artifact of a filled-in BamHI
FT      site; the Gly separating the Region I and II contg. CS flanking
FT      regions is an artifact of a synthetic FokI/ThiIII I linker. The
FT      peptide can be used in a vaccine for protection against malaria.
FT      See also AAR12306-R12311 and AAR13175-R13179.
FT
FT      Sequence 335 AA;
FT
FT      Query Match          100.0%; Score 109; DB 12; Length 335;
FT      Best Local Similarity 100.0%; Pred. No. 7.7e-09;
FT      Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY      1 DIEKICKMEKCSVFNVVNS 21
        ||||||||||||||||
DB      301 diekickmekcsvfnnvns 321

RESULT  7
AAR13179
ID      AAR13179 standard; Protein: 335 AA.
XX
AC      AAR13179;
XX
DT      29-AUG-1991 (first entry)
XX
DE      NS1_81(NVDP)4RLfAuth.
XX
KM      Immunogenic determinant; circumsporozoite; CS; vaccine; malaria;
XX      hybrid; Influenza virus; non-structural protein 1; fusion.
XX
OS      Plasmodium falciparum.
OS      Influenza virus (A/PR/8/34).
XX
FH      Key
FT      Location/Qualifiers
FT      1..81
FT      /label= N-terminal of NS1
FT      /note= "Influenza virus nonstructural protein 1"
FT      82..97
FT      /label= Immunodominant repeat region
FT      /note= "four variant tetrapeptide repeat units"
FT      98..103
FT      /label= synthetic linker
FT      104
FT      /label= artifact
FT      /note= "see comments"
FT      105..209
FT      /label= Aas 19-123 of CS protein
FT      /note= "Region I contg. flanking region less
FT      signal sequence"
FT      210
FT      /label= artifact
FT      /note= "see comments"
FT      211..335
FT      /label= Aas 288-412 of CS protein
FT      /note= "Region II flanking region"
FT      Region
FT      19-JUN-1991.
PD      EP432965-A.
XX
PN      19-JUN-1991.
XX
PF      06-DEC-1990; 90BP-0313257.
XX
PR      08-DEC-1989; 89US-0447746.
XX
PA      (SMK ) SMITHKLINE BEECHAM.
PA      (USSA ) US SEC OF THE ARMY.
PA      (BIOM-) BIOMEDICAL RES INST.
XX
PI      Gross MS, Gordon DM, Hollingdale MR.
XX      WPI: 1991-179771/25.
DR      WPI: 1991-179771/25.
XX
PT      Polyptide comprising immunogenic determinants from P falciparum
XX      - for vaccine against malaria infection in humans.
XX
PS      Example 5; Page 11; 18pp; English.
XX
CC      The polyptide is prepd. by genetic engineering of genes encoding
CC      the P. falciparum circumsporozoite (CS) protein [Dame et al.,
CC      Science 225 : 593 (1984)], and the influenza virus non-structural
CC      protein 1 (NS1), [Baez et al., Nucleic Acids Research, 8 : 5845
CC      (1980)]. The DNA encoding the 1st 81 AAs of the N-terminal of NS1
CC      (NS1_81) is linked to a synthetic sequence encoding four repeat
CC      units (the variant form) from the immunodominant region, which in
CC      turn is linked via a synthetic sequence to DNA encoding Region I

```

```

CC      contg. flanking region less the 18 AA signal region. This is
CC      linked to DNA encoding Region II-contg. flanking region. The pro
CC      residue separating the Asp (at the C-terminal of the linker) from
CC      the Region I-contg. CS flanking region is an artifact of a filled-
CC      in BamHI site; the Gly separating the Region I and II-contg. CS
CC      flanking regions is an artifact of a synthetic FokI/NotIII I
CC      linker. The peptide can be used in a vaccine for protection
CC      against malaria.
CC      See also AAR12306-R12311 and AAR13175-R13178.
XX
SQ      Sequence      335 AA:
XX
XX
Query Match      100.0%; Score 109; DB 12; Length 335;
Best Local Similarity 100.0%; Pred. No. 7.7e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 DIEKICKMEKCSVFNVVNS 21
        ||||||||||||||||
DB      301 diekickmekcsvfnnvns 321

RESULT  8
AAY70278
ID      AAY70278 standard; Protein: 350 AA.
XX
AC      AAY70278;
XX
DT      06-JUN-2000 (first entry)
XX
DE      Recombinant vaccine CDC/NIIMALVAC-1.
XX
KM      Recombinant protein; CDC/NIIMALVAC-1; multivalent; malaria; vaccine;
KM      T-cell epitope; tetanus toxoid; antigenic epitope; treatment;
KM      circumsporozoite protein; CSP; sporozoite surface protein-2; SSP-2;
KM      liver stage antigen-1; LSA-1; merozoite surface protein-1; MSP-1; MSP-2;
KM      apical membrane antigen-1; AMA-1; erythrocyte binding antigen-175;
KM      Pfgr27; antiparasitic; prevention; anti-CDC/NIIMALVAC-1 antibody;
KM      honey bee.
XX
OS      Chimeric - Apis sp.
OS      Chimeric - Clostridium tetani.
OS      Chimeric - Plasmodium falciparum.
XX
FH      Key
FT      Location/Qualifiers
FT      1..22
FT      /label= Melittin_signal_peptide
FT      /note= "Derived from Honey bee"
FT      23..350
FT      /label= Mature_CDC/NIIMALVAC-1
FT      /note= "Recombinant multivalent malarial vaccine"
XX
PD      WO200011179-A1.
XX
PN      02-MAR-2000.
XX
PF      19-AUG-1999; 99WO-US18869.
XX
PR      21-AUG-1998; 98US-0097703.
XX
PA      (NATIM-) NAT INST. IMMUNOLOGY.
PA      (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI      Lal AA, Shi YP, Hasnain SE;
XX      WPI: 2000-237654/20.
XX      N-PsDB; AA251336.
XX
PT      Novel recombinant protein as vaccine for treating malarial infection
XX      comprises antigenic peptides obtained from different stages of
XX      plasmodium falciparum life cycle

```

PS Claim 3; Page 43-44; 52pp; English.  
 XX The present sequence is that of recombinant protein CDC/NIIMALVAC-1,  
 CC which is a multivalent, multistage malarial vaccine. The recombinant  
 CC protein comprises, melittin signal peptide, (His)6 tag, T-cell epitope  
 CC from tetanus toxoid and 21 antigenic epitopes from circumsporozoite  
 CC protein (CSP), sporozoite surface protein-1 (MSP-1), MSP-2, apical  
 CC antigen-1 (LSA-1), merozoite surface binding antigen-175 (EBA-175),  
 CC membrane antigen-1 (AMA-1), erythrocyte binding antigen, PfG27.  
 CC Thoptry associated protein-1 (RAP-1) and gamete specific antigen, PfG27.  
 CC These epitopes were obtained at different stages of the life cycle of  
 CC Plasmodium falciparum. CDC/NIIMALVAC-1 vaccine has antiparasitic  
 CC activity and can used for treatment and prevention of malarial  
 CC infections. Anti-CDC/NIIMALVAC-1 antibodies can be used for detecting  
 CC P. falciparum in biological samples.  
 CC  
 SQ Sequence 350 AA;

Query Match 100.0%; Score 109; DB 21; Length 350;  
 Best Local Similarity 100.0%; Pred. No. 8,1e-09;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DIEKICKMEKCSVFNNVNS 21  
 ||||||||||||||||  
 DB 61 diekikmekcsvfnnvns 81

RESULT 9  
 AAP83144  
 ID AAP83144 standard; protein; 411 AA.

XX AAP83144;  
 XX 20-NOV-1990 (first entry)  
 XX  
 XX Sequence encoded by the circumsporozoite (CS) gene from  
 DE Plasmodium falciparum.  
 XX  
 XX Vaccine; antigen; immunogen; probe; hybridisation;  
 KW immunosassay; diagnosis.  
 XX  
 XX Plasmodium falciparum.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH 106..120  
 FT Region /note="Region 1"  
 FT 147..206  
 FT Region /note="Repeat region, repeat unit = NAMP"  
 FT 123..146  
 FT Region /note="Repeat region, repeat unit = NAMPNVPD"  
 FT 211..286  
 FT Region /note="Repeat region, repeat unit = NAMP"  
 XX  
 XX EP278940-A.  
 PN  
 XX 17-AUG-1988.  
 PD  
 XX 25-JAN-1988; 88EP-0870008.  
 PE  
 XX 30-JAN-1987; 87US-0009325.  
 PR  
 XX (SMIK ) SMITH KLINE-RIT.  
 PA  
 XX Cabezon T, De Wilde M, Harford N;  
 PI  
 XX WPI: 1988-229751/33.  
 DR  
 XX N-PSDB; AAN81108.  
 DR  
 XX DNA encoding hepatitis B virus antigens and hybrids contg. them  
 PT used for expression in yeast to obtain vaccines and bivalent  
 PT vaccines  
 XX

PS Example; Fig 3Aa-3Af; 101pp; English.

XX Sequence of the CS gene (AAN81108) is from lambda-mpf1. A recombinant  
 CC DNA molecule is claimed, comprising functional DNA coding sequence fused,  
 CC in phase, to a portion of the Pre-S2 region of a hepatitis B virus (HBV)  
 CC Pre-S2-S protein coding sequence. The functional DNA coding sequence of  
 CC comprises the Pre-S2 coding sequence, the CS protein coding sequence of  
 CC Pre-S1-Pre-S2 coding sequence, the CS protein coding sequence of  
 CC Plasmodium, or a HIV coding sequence such as  
 CC an HIV envelope gene sequence, e.g., HIV C7 protein coding region, HIV  
 CC Peptide 121 coding region, or HIV Dreesman peptide coding region.  
 CC  
 SQ Sequence 411 AA;

Query Match 100.0%; Score 109; DB 9; Length 411;  
 Best Local Similarity 100.0%; Pred. No. 9,6e-09;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DIEKICKMEKCSVFNNVNS 21  
 ||||||||||||||||  
 DB 377 diekikmekcsvfnnvns 397

RESULT 10  
 AAP60416  
 ID AAP60416 standard; protein; 412 AA.

XX AAP60416;  
 XX 13-JUN-1991 (first entry)  
 XX  
 XX CS protein of malaria parasite.  
 DE  
 XX Sporozoite; vaccination.  
 KW  
 XX Plasmodium falciparum.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH 124..127  
 FT Region /label= Repeat unit  
 FT  
 FT  
 FT  
 XX  
 XX EP16410-A.  
 PN  
 XX 02-JAN-1986.  
 PD  
 XX 24-JUN-1985; 85EP-0107794.  
 PE  
 XX 26-JUN-1984; 84US-0624564.  
 PR  
 XX (USDC ) US SEC OF COMMERCE.  
 PA (USGO ) US GOVERNMENT.  
 PA (USSA ) US SEC OF THE ARMY.  
 XX  
 XX McCutchan TF, Dame JB, Williams JL, Schneider I;  
 PI  
 XX WPI: 1986-008635/02.  
 DR N-PSDB; AAN60362.  
 DR  
 XX New immunologically active pure synthetic peptide(s) - used for  
 PT protection against infection by malaria parasite.  
 PT  
 XX Disclosure; Fig 2; 49pp; English.  
 PS  
 XX The Plasmodium CS gene was used to isolate peptides capable of  
 CC inducing an immune response to the parasite. Peptide antigens may  
 CC be synthesised in pure form and used to generate an immune  
 CC response in vaccination against malaria. The featured repeat  
 CC units are claimed and must be present in copies of 2-1000.  
 CC  
 SQ Sequence 412 AA;



Query Match 100.0%; Score 109; DB 7; Length 412;  
 Best Local Similarity 100.0%; Pred. No. 9.6e-09;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DIEKICKMEKCSVFNVNS 21  
 ||||||||||||||||  
 DB 378 diekikmekcsvfnnvns 398

## RESULT 11

AAK37796  
 ID AAK37796 standard; Protein: 424 AA.

AC AAK37796;

DT 27-SEP-1993 (first entry)

DE RTS protein.

KW RTS; expression cassette: hybrid protein: S. cerevisiae; TDH3;  
 cloning; circumsporozoite protein; CSP; Plasmodium falciparum;  
 KW strain 7G8; hepatitis B virus; HBV; adw serotype; pres2 protein;  
 S protein.

OS Synthetic.

XX Location/Qualifiers

FT 1 /note= "Derived from S. cerevisiae TDH3 gene sequence"

FT 2..4 /note= "Cloning artefact"

FT 5..193 /note= "Represents amino acids 210-398 of the CSP of

FT 194..197 P. falciparum"

FT /note= "Carboxy terminal amino acids from HBV (adw

FT 198..424 serotype) pres2 protein"

FT /note= "S protein of HBV (adw serotype)"

XX W09310152-A.

PD 27-MAY-1993.

PE 11-NOV-1992; 92WO-EP02591.

PR 16-NOV-1991; 91GB-0024390.

PR 27-FEB-1992; 92US-0842694.

XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

PI Cohen J, De Wilde M;

DR WPI: 1993-182494/22.

DR N-PSDB: AA042566.

XX Hybrid protein comprising Plasmodium circumsporozoite protein and

PT HBsAg - useful as a vaccine for treating patients susceptible to

PT Plasmodium infections

XX Disclosure: Fig 5; 59pp; English.

XX This sequence represents the RTS hybrid protein which is encoded by  
 CC the RTS expression cassette. This hybrid consists of a methionine  
 CC residue derived from S. cerevisiae TDH3 gene sequence, three amino  
 CC acids, Met-Ala-Pro, derived from a nucleotide sequence created by  
 CC the cloning procedure used to construct the hybrid gene, a stretch  
 CC of 189 amino acids representing amino acids 210 to 398 of the  
 CC circumsporozoite protein (CSP) of Plasmodium falciparum strain 7G8,  
 CC an amino acid Arg created by the cloning procedure, four amino acids,  
 CC Pro-Val-Thr-Asn, representing the four carboxy terminal residues of  
 CC hepatitis B virus (HBV), adw serotype, pres2 protein, and a stretch  
 CC of 226 amino acids specifying the S protein of HBV, adw serotype.

CC This protein, and RTS\* (see also AAK37797), may be combined with an  
 CC adjuvant and used in a vaccine for preventing plasmodium infections.  
 CC The vaccines produce a humoral response and also a cellular immune  
 CC response.

XX Sequence 424 AA:

Query Match 100.0%; Score 109; DB 14; Length 424;  
 Best Local Similarity 100.0%; Pred. No. 9.9e-09;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DIEKICKMEKCSVFNVNS 21  
 ||||||||||||||||  
 DB 173 diekikmekcsvfnnvns 193

## RESULT 12

AAK37797  
 ID AAK37797 standard; Protein: 424 AA.

AC AAK37797;

DT 27-SEP-1993 (first entry)

DE RTS\* protein.

KW RTS; expression cassette: hybrid protein: S. cerevisiae; TDH3;  
 cloning; circumsporozoite protein; CSP; Plasmodium falciparum;  
 KW strain 7G8; hepatitis B virus; HBV; adw serotype; pres2 protein;  
 S protein.

OS Synthetic.

XX Location/Qualifiers

FT 1 /note= "Derived from S. cerevisiae TDH3 gene sequence"

FT 2..4 /note= "Cloning artefact"

FT 5..193 /note= "Represents amino acids 210-398 of the CSP of

FT 194..197 P. falciparum"

FT /note= "Carboxy terminal amino acids from HBV (adw

FT 198..424 serotype) pres2 protein"

FT /note= "S protein of HBV (adw serotype)"

XX W09310152-A.

PD 27-MAY-1993.

PE 11-NOV-1992; 92WO-EP02591.

PR 16-NOV-1991; 91GB-0024390.

PR 27-FEB-1992; 92US-0842694.

XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

PI Cohen J, De Wilde M;

DR WPI: 1993-182494/22.

DR N-PSDB: AA042567.

XX Hybrid protein comprising Plasmodium circumsporozoite protein and

PT HBsAg - useful as a vaccine for treating patients susceptible to

PT Plasmodium infections

XX Disclosure: Fig 9; 59pp; English.

XX This sequence represents the RTS\* hybrid protein which is encoded by  
 CC the RTS\* expression cassette. This hybrid consists of a methionine  
 CC residue derived from S. cerevisiae TDH3 gene sequence, three amino



CC of the native protein's Cys residues at positions 384 and 389. Also  
 CC claimed is AAP91504 (or modified forms, see PT) associated with an  
 CC antigenic structure representing a B-cell epitope, pref. a multiple  
 CC antigenic peptide, esp. multimers of the repeat sequences NAMP  
 CC present in P. falciparum CS protein.  
 XX  
 SQ Sequence 21 AA:

Query Match 83.5%; Score 91; DB 10; Length 21;  
 Best Local Similarity 90.5%; Pred. No. 2.4e-07;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 DIEKKICKMEKCSVFNVNS 21  
 ||||| |||| |||||  
 Db 1 diekkikmekesvfnvns 21

RESULT 15

AAR82586  
 ID AAR82586 standard; peptide: 21 AA.

XX  
 AC AAR82586;

XX  
 DT 13-JUN-1996 (first entry)

XX  
 DE Plasmodium falciparum circumsporozoite helper T cell epitope, PF.

XX  
 KW IGE; CH4; Immunoglobulin; epsilon; Immunogen; helper T cell; epitope;  
 KW vaccine; allergy; antibody; constant heavy chain.

XX  
 OS Plasmodium falciparum.

XX  
 PN W09526365-A1.

XX  
 PD 05-OCT-1995.

XX  
 PF 24-MAR-1995; 95WO-US03741.

XX  
 PR 25-OCT-1994; 94US-0328912.

XX  
 PR 28-MAR-1994; 94US-0218461.

XX  
 PA (UNBI-) UNITED BIOMEDICAL INC.

XX  
 PI Wang CY;

XX  
 DR WPI; 1995-351297/45.

XX  
 PT Synthetic peptide-based immunogen contg. IGE CH4 peptide and helper  
 PT T cell epitope - useful for eliciting antibody prodn. for allergy  
 PT treatment

XX  
 PS Claim 3; Page 23; 87pp; English.

XX  
 CC AAR82571-91 are helper T cell epitopes which can be used in the  
 CC preparation of a peptide immunogen that is useful in vaccines for  
 CC treating allergic reactions. In the immunogen an IGE CH4 peptide  
 CC is attached C-terminally to a series of amino acids including a  
 CC helper T cell epitope. The immunogen may also opt. contain a fatty  
 CC acid or fatty acid derivative, an invasins domain or alpha-NH2. The  
 CC immunogen produces high titres of antibodies to the effector site  
 CC in human IGE heavy chain (the CH4 domain peptide) which inhibit mast  
 CC cell activation and reduce allergen-induced IGE prodn. The immunogens  
 CC may be used in either a radially branching multimeric form or a  
 CC linearly arranged monomeric form.

XX  
 SQ Sequence 21 AA:

Query Match 83.5%; Score 91; DB 16; Length 21;  
 Best Local Similarity 90.5%; Pred. No. 2.4e-07;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 DIEKKICKMEKCSVFNVNS 21  
 ||||| |||| |||||  
 Db 1 diekkikmekesvfnvns 21

Search completed: January 29, 2002, 10:21:43  
 Job time: 419 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using SW model

Run on: January 29, 2002, 10:24:03 ; Search time 133.18 Seconds  
(without alignments)  
3.548 Million cell updates/sec

Title: US-09-763-397A-7  
Perfect score: 109  
Sequence: 1 DIEKICKMEKSSVFNVNS 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Dackfilest1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	109	100.0	412	1	US-08-313-288B-18
2	109	100.0	423	2	US-08-760-797A-1
3	109	100.0	424	2	US-08-760-797A-3
4	109	100.0	424	4	US-08-932-929B-1
5	109	100.0	424	4	US-08-932-929B-3
6	91	83.5	21	1	US-08-186-266-6
7	91	83.5	21	1	US-08-446-692-48
8	91	83.5	21	2	US-08-488-351A-48
9	91	83.5	21	3	US-09-100-409A-54
10	91	83.5	21	5	PCR-US95-02121-97
11	91	83.5	21	5	PCR-US95-13841-20
12	91	83.5	33	1	US-08-446-692-27
13	91	83.5	33	2	US-08-488-351A-27
14	83	76.1	478	3	US-08-155-888-2
15	75.5	69.3	20	1	US-08-465-167A-20
16	75.5	69.3	20	5	PCR-US92-07218-17
17	71	65.1	15	2	US-08-318-856A-75
18	71	65.1	17	5	PCR-US95-02121-96
19	71	65.1	66	2	US-08-455-625-35
20	71	65.1	66	4	US-08-455-685-35
21	71	65.1	66	4	US-08-060-988A-35
22	71	65.1	66	5	PCR-US94-05142-35
23	69	63.3	378	6	5171843-9
24	67	61.5	16	2	US-08-817-933A-7
25	62	56.9	15	6	5169933-30
26	61	56.0	21	1	US-08-305-871A-12
27	55.5	50.9	16	1	US-08-465-167A-19

28	55.5	50.9	16	5	PCT-US92-07218-16	Sequence 16, Appl
29	48	44.0	15	2	US-08-318-856A-76	Sequence 76, Appl
30	47	43.1	492	4	US-08-845-258-21	Sequence 21, Appl
31	47	43.1	492	4	US-08-990-571-21	Sequence 21, Appl
32	47	43.1	492	4	US-08-723-142A-21	Sequence 21, Appl
33	47	43.1	503	4	US-08-845-258-52	Sequence 52, Appl
34	47	43.1	503	4	US-08-990-571-52	Sequence 52, Appl
35	46	42.2	1400	1	US-08-080-255-7	Sequence 7, Appl
36	46	42.2	1400	3	US-08-465-713-7	Sequence 7, Appl
37	46	42.2	1400	5	PCT-US93-05857-7	Sequence 7, Appl
38	46	42.2	3969	4	US-08-061-376-5	Sequence 5, Appl
39	39.5	36.2	159	6	5208144-35	Patent No. 5208144
40	39	35.8	77	2	US-08-465-380-58	Sequence 58, Appl
41	39	35.8	77	2	US-08-486-397-58	Sequence 58, Appl
42	39	35.8	77	2	US-08-486-399-58	Sequence 58, Appl
43	39	35.8	77	2	US-08-461-965-58	Sequence 58, Appl
44	39	35.8	77	2	US-08-634-641-58	Sequence 58, Appl
45	39	35.8	77	3	US-09-249-471-58	Sequence 58, Appl

## ALIGNMENTS

RESULT 1  
US-08-313-288B-18  
Sequence 18, Application US/08313288B  
Patent No. 5750502  
GENERAL INFORMATION:  
APPLICANT: Jessell, Thomas M. and Avihu Klar  
TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/313,288B  
FILING DATE: January 5, 1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 40028-A-PCT-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0526  
TELEX:  
INFORMATION FOR SEQ. ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 412 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-313-288B-18  
Query Match 100.0%; Score 109; DB 1; Length 412;  
Best local Similarity 100.0%; Pred. No. 8.4e-09;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 DIEKICKMEKSSVFNVNS 21  
DB 378 DIEKICKMEKSSVFNVNS 398

RESULT 2  
US-08-760-797A-1  
; Sequence 1, Application US/08760797A  
; Patent No. 5928902  
; GENERAL INFORMATION:  
; APPLICANT: De Wilde, Michel  
; APPLICANT: Cohen, Joseph  
; TITLE OF INVENTION: Hybrid Protein Between CS  
; TITLE OF INVENTION: from Plasmodium and HBSAG  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/760,797A  
; FILING DATE: 04-DEC-1996  
; CLASSIFICATION: 435  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: 08/442,612  
; FILING DATE: 17-MAY-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Baumeister, Kirk  
; REGISTRATION NUMBER: 33,833  
; REFERENCE/DOCKET NUMBER: B45015-1C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-270-5096  
; TELEFAX: 610-270-5090  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 423 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-760-797A-1

Query Match 100.0%; Score 109; DB 2; Length 423;  
Best Local Similarity 100.0%; Pred. No. 8,6e-09;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DIEKICKMEKCSYFNVNVS 21  
DB 172 DIEKICKMEKCSYFNVNVS 192

RESULT 3  
US-08-760-797A-3  
; Sequence 3, Application US/08760797A  
; Patent No. 5928902  
; GENERAL INFORMATION:  
; APPLICANT: De Wilde, Michel  
; APPLICANT: Cohen, Joseph  
; TITLE OF INVENTION: Hybrid Protein Between CS  
; TITLE OF INVENTION: from Plasmodium and HBSAG  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA

ZIP: 19406  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/760,797A  
; FILING DATE: 04-DEC-1996  
; CLASSIFICATION: 435  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: 08/442,612  
; FILING DATE: 17-MAY-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Baumeister, Kirk  
; REGISTRATION NUMBER: 33,833  
; REFERENCE/DOCKET NUMBER: B45015-1C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-270-5096  
; TELEFAX: 610-270-5090  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 424 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-760-797A-3

Query Match 100.0%; Score 109; DB 2; Length 424;  
Best Local Similarity 100.0%; Pred. No. 8,7e-09;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DIEKICKMEKCSYFNVNVS 21  
DB 169 DIEKICKMEKCSYFNVNVS 189

RESULT 4  
US-08-932-929B-1  
; Sequence 1, Application US/08932929B  
; Patent No. 6169171  
; GENERAL INFORMATION:  
; APPLICANT: De Wilde, Michel  
; APPLICANT: Cohen, Joseph  
; TITLE OF INVENTION: Hybrid Protein Between CS  
; TITLE OF INVENTION: from Plasmodium and HBSAG  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/932,929B  
; FILING DATE:  
; CLASSIFICATION:  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: 08/760,797  
; FILING DATE: 04-DEC-1996  
; APPLICATION NUMBER: 08/442,612  
; FILING DATE: 17-MAY-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Baumeister, Kirk  
; REGISTRATION NUMBER: 33,833

```
REFERENCE/DOCKET NUMBER: B45015-1FWC2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5096
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 424 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-932-9298-1

Query Match      100.0%; Score 109; DB 4; Length 424;
Best Local Similarity 100.0%; Pred. No. 8.7e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DIEKKICKMEKSSVFNVVNS 21
    |||||
Db 173 DIEKKICKMEKSSVFNVVNS 193

RESULT 5
US-08-932-9298-3
; Sequence 3, Application US/08932929B
; Patent No. 6169171
; GENERAL INFORMATION:
; APPLICANT: De Wilde, Michel
; APPLICANT: Cohen, Joseph
; TITLE OF INVENTION: Hybrid Protein Between CS
; TITLE OF INVENTION: from Plasmodium and HbsAG
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Smithkline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/932,929B
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/760,797
; FILING DATE: 04-DEC-1996
; APPLICATION NUMBER: 08/442,612
; FILING DATE: 17-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Baumeister, Kirk
; REGISTRATION NUMBER: 33,833
; REFERENCE/DOCKET NUMBER: B45015-1FWC2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5096
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 424 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-932-9298-3

Query Match      100.0%; Score 109; DB 4; Length 424;
Best Local Similarity 100.0%; Pred. No. 8.7e-09;
```

```
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DIEKKICKMEKSSVFNVVNS 21
    |||||
Db 169 DIEKKICKMEKSSVFNVVNS 189

RESULT 6
US-08-186-266-6
; Sequence 6, Application US/08186266
; Patent No. 5662907
; GENERAL INFORMATION:
; APPLICANT: KUBO, Ralph T.
; APPLICANT: GREY, Howard M.
; APPLICANT: SETTE, Alessandro
; APPLICANT: CELIS, Esteban
; TITLE OF INVENTION: INDUCTION OF ANTI-TUMOR CYTOTOXIC
; TITLE OF INVENTION: T LYMPHOCYTES IN HUMANS USING
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: Stewart Street Tower, One Market Plaza
; CITY: San Francisco
; STATE: California
; COUNTRY: US
; ZIP: 94105-1493
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/186,266
; FILING DATE: 25-JAN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/159,339
; FILING DATE: 29-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/103,396
; FILING DATE: 06-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/027,746
; FILING DATE: 05-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/926,666
; FILING DATE: 07-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 14137-50-4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..21
; OTHER INFORMATION: /note="Plasmodium falciparum CS
; OTHER INFORMATION: protein at positions 378-398."
US-08-186-266-6

Query Match      83.5%; Score 91; DB 1; Length 21;
Best Local Similarity 90.5%; Pred. No. 2.1e-07;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 DIEKICKMEKCSSVFNVNS 21
    ||||| |||| |
Db 1 DIEKIAMEKASSVFNVNS 21
```

```

1      RESULT 7
2      US-08-446-692-48
3      : Sequence 48, Application US/08446692
4      : Patent No. 5759551
5      :
6      : GENERAL INFORMATION:
7      :
8      : APPLICANT: Ladd, Anna
9      : APPLICANT: Wang, Chang Yi
10     : APPLICANT: Zamb, Timothy
11     : TITLE OF INVENTION: Immunogenic LHRH peptide constructs
12     : TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines
13     : NUMBER OF SEQUENCES: 114
14     : CORRESPONDENCE ADDRESS:
15     : ADDRESSEE: Maria C.H. Lin
16     : STREET: 345 Park Avenue
17     : CITY: New York
18     : STATE: NY
19     : COUNTRY: US
20     : ZIP: 10154-0053
21     : COMPUTER READABLE FORM:
22     : MEDIUM TYPE: Floppy disk
23     : COMPUTER: IBM PC compatible
24     : OPERATING SYSTEM: PC-DOS/MS-DOS
25     : SOFTWARE: Patentin Release #1.0, Version #1.25
26     :
27     : CURRENT APPLICATION DATA:
28     : APPLICATION NUMBER: US/08/446,692
29     : FILING DATE: 7-JUN-1995
30     : CLASSIFICATION: 424
31     : ATTORNEY/AGENT INFORMATION:
32     : NAME: Maria C.H. Lin
33     : REGISTRATION NUMBER: 29,323
34     : REFERENCE/DOCKET NUMBER: 1151-4146 US2
35     : TELECOMMUNICATION INFORMATION:
36     : TELEPHONE: (212)415-8745
37     : TELEFAX: (516)751-6849
38     : INFORMATION FOR SEQ ID NO: 48:
39     : SEQUENCE CHARACTERISTICS:
40     : LENGTH: 21 amino acids
41     : TYPE: amino acid
42     : TOPOLOGY: linear
43     :
44     : MOLECULE TYPE: peptide
45     :
46     : US-08-446-692-48

```

```

Query Match      83.5%; Score 91; DB 1; Length 21;
Best Local Similarity 90.5%; Pred. No. 2.1e-07;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      1 DIEKRICKMEKCSVFNVNVS 21
        ||||| ||| |||||
Db      1 DIEKRIAKMEKASSVFNVNS 21

RESULT      8
US-08-488-351A-48
; Sequence 48, Application US/08488351A
; Patent No. 5843446
;
; GENERAL INFORMATION:
; APPLICANT: Ladd, Anna
; APPLICANT: Wang, Chang YI
; APPLICANT: Zamb, Timothy
; TITLE OF INVENTION: Immunogenic LHRH peptide constructs
; and synthetic universal immune stimulants for vaccines
;
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Maria C.H. Lin
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
;
;
;

```

```

COUNTRY: US
ZIP: 10154-0053
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,351A
FILING DATE: 7-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/446,692
FILING DATE: 7-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/229,275
FILING DATE: 14-APR-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/057,166
FILING DATE: 27-APR-1992
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maria C.H. Lin
REGISTRATION NUMBER: 29,323
REFERENCE/DOCKET NUMBER: 1151-4146 US2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)415-8745
TELEFAX: (516)751-6649
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-488-351A-48

```

```

Query Match          83.5%;      Score 91;  DB 2;  Length 21;
Best Local Similarity 90.5%;      Pred. NO. 2.1e-07;
Matches 19;  Conservative 0;  Mismatches 2;  Indels 0;  Gaps 0;

OY      1  DIEKKICKMEKCSVFENVVNS 21
        ||||| ||||| ||||| |||
DB      1  DIEKKIAMEKASVFENVVNS 21

RESULT  9
US-09-100-409A-54
; Sequence 54, Application US/09100409A
; Patent No. 6090388
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang YI
; TITLE OF INVENTION: PEPTIDE COMPOSITION FOR
; TITLE OF INVENTION: PREVENTION AND TREATMENT OF HIV INFECTION AND
; TITLE OF INVENTION: IMMUNE DISORDERS
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10154-0054
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patentin Release #1.0, Version
; SOFTWARE: #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/100.409A

```



FILED DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME:  
REGISTRATION NUMBER:  
REFERENCE/DOCKET NUMBER: 1151-4154  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-758-4800  
TELEFAX: 212-751-6849  
INFORMATION FOR SEQ ID NO: 54:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-100-409A-54

Query Match 83.5%; Score 91; DB 3; Length 21;  
Best Local Similarity 90.5%; Pred. No. 2.1e-07;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 DIEKICKMEKSSVFNVVNS 21  
||||| ||||| ||||| ||||| |||||  
Db 1 DIEKIKAKMEKASSVFNVVNS 21

RESULT 10  
PCT-US95-02121-97  
Sequence 97, Application PC/TUS9502121  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR ELICITING  
TITLE OF INVENTION: CTL IMMUNITY  
NUMBER OF SEQUENCES: 153  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/02121  
FILING DATE: 16-FEB-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/197,484  
FILING DATE: 16-FEB-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/935,811  
FILING DATE: 26-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/874,491  
FILING DATE: 27-APR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/827,682  
FILING DATE: 29-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/749,568  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Parmelee, Steven W.  
REGISTRATION NUMBER: 31,990  
REFERENCE/DOCKET NUMBER: 14137-26-4PC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 467-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 97:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide

FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..21  
OTHER INFORMATION: /note="Malaria circumsporozoite"  
OTHER INFORMATION: 378-398"  
PCT-US95-02121-97

Query Match 83.5%; Score 91; DB 5; Length 21;  
Best Local Similarity 90.5%; Pred. No. 2.1e-07;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 DIEKICKMEKSSVFNVVNS 21  
||||| ||||| ||||| ||||| |||||  
Db 1 DIEKIKAKMEKASSVFNVVNS 21

RESULT 11  
PCT-US95-13841-20  
Sequence 20, Application PC/TUS9513841  
GENERAL INFORMATION:  
APPLICANT: United Biomedical Inc; Walfield, Alan M.;  
APPLICANT: Wang, Chang Yi  
TITLE OF INVENTION: Synthetic Ige Membrane Anchor  
TITLE OF INVENTION: Peptide Immunogens for the Treatment of Allergy  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Maria C.H. Lin  
STREET: 345 Park Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/13841  
FILING DATE: 25-OCT-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/328,519  
FILING DATE: 25-OCT-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Lin, Maria C.H.  
REGISTRATION NUMBER: 29,323  
REFERENCE/DOCKET NUMBER: 1151-4117  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-758-4800  
TELEFAX: 212-751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US95-13841-20

Query Match 83.5%; Score 91; DB 5; Length 21;  
Best Local Similarity 90.5%; Pred. No. 2.1e-07;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 DIEKICKMEKSSVFNVVNS 21  
||||| ||||| ||||| ||||| |||||  
Db 1 DIEKIKAKMEKASSVFNVVNS 21

RESULT 12

US-08-446-692-27  
; Sequence 27, Application US/08446692  
; Patent No. 5759551  
; GENERAL INFORMATION:  
; APPLICANT: Ladd, Anna  
; APPLICANT: Wang, Chang Yi  
; APPLICANT: Zamb, Timothy  
; TITLE OF INVENTION: Immunogenic LHRH peptide constructs  
; TITLE OF INVENTION: and synthetic universal immune stimulants for vaccines  
; NUMBER OF SEQUENCES: 114  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Maria C.H. Lin  
; STREET: 345 Park Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: US  
; ZIP: 10154-0053  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/446,692  
; FILING DATE: 7-JUN-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maria C.H. Lin  
; REGISTRATION NUMBER: 29,323  
; REFERENCE/DOCKET NUMBER: 1151-4146 US2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)415-8745  
; TELEFAX: (516)751-6849  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 33 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-446-692-27

Query Match 83.5%; Score 91; DB 1; Length 33;  
Best Local Similarity 90.5%; Pred. No. 3.3e-07;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 DIEKKICKMEKCSVFNVNS 21  
Db 1 DIEKKIAKMEKASSVFNVNS 21

RESULT 13  
US-08-488-351A-27  
; Sequence 27, Application US/08488351A  
; Patent No. 5843446  
; GENERAL INFORMATION:  
; APPLICANT: Ladd, Anna  
; APPLICANT: Wang, Chang Yi  
; APPLICANT: Zamb, Timothy  
; TITLE OF INVENTION: Immunogenic LHRH peptide constructs  
; TITLE OF INVENTION: and synthetic universal immune stimulants for vaccines  
; NUMBER OF SEQUENCES: 114  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Maria C.H. Lin  
; STREET: 345 Park Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: US  
; ZIP: 10154-0053  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/488,351A  
; FILING DATE: 7-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/446,692  
; FILING DATE: 7-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/229,275  
; FILING DATE: 14-APR-1994  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/057,166  
; FILING DATE: 27-APR-1992  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maria C.H. Lin  
; REGISTRATION NUMBER: 29,323  
; REFERENCE/DOCKET NUMBER: 1151-4146 US2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)415-8745  
; TELEFAX: (516)751-6849  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 33 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-488-351A-27

Query Match 83.5%; Score 91; DB 2; Length 33;  
Best Local Similarity 90.5%; Pred. No. 3.3e-07;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 DIEKKICKMEKCSVFNVNS 21  
Db 1 DIEKKIAKMEKASSVFNVNS 21

RESULT 14  
US-08-155-888-2  
; Sequence 2, Application US/08155888  
; Patent No. 6066623  
; GENERAL INFORMATION:  
; APPLICANT: Hoffman, Stephen L.  
; APPLICANT: Hedstrom, Richard C.  
; APPLICANT: Sedegah, Martha  
; TITLE OF INVENTION: POLYNUCLEOTIDE VACCINE PROTECTIVE  
; TITLE OF INVENTION: AGAINST MALARIA, METHODS OF PROTECTION AND VECTOR FOR  
; TITLE OF INVENTION: DELIVERING POLYNUCLEOTIDE VACCINES  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Naval Medical Res. & Dev. Cnd.  
; STREET: Bldg. 1, T-12 8901 Wisconsin Ave.  
; CITY: Bethesda  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20889-5606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/155,888  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Spevack, A. David  
; REGISTRATION NUMBER: 24,743

REFERENCE/DOCKET NUMBER: N.C. 75,851  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 295-6759  
 TELEFAX: (202) 295-1022  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 478 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-155-888-2

Query Match 76.1%; Score 83; DB 3; Length 478;  
 Best Local Similarity 61.9%; Pred. NO. 6.6e-05;  
 Matches 13; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

OY 1 DIEKKICKMEKCSYFNVVNS 21  
 ||:|||||:||||:|:  
 Db 444 DIDFKKMKCSIFNIVSN 464

RESULT 15  
 US-08-465-167A-20  
 ; Sequence 20, Application US/08465167A  
 ; Patent No. 5750395  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fikes, John D.  
 ; APPLICANT: Livingston, Brian D.  
 ; APPLICANT: Sette, Alessandro D.  
 ; APPLICANT: Sidney, John C.  
 ; TITLE OF INVENTION: DNA ENCODING MAGE-1 C-TERMINAL  
 ; TITLE OF INVENTION: IMMUNOGENIC PEPTIDES (as amended)  
 ; NUMBER OF SEQUENCES: 51  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Townsend and Townsend and Crew LLP  
 ; STREET: Two Embarcadero Center, 8th floor  
 ; CITY: San Francisco  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 96111  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/465,167A  
 ; FILING DATE: 05-JUN-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/103,623  
 ; FILING DATE: 06-AUG-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Parmelee, Steven W.  
 ; REGISTRATION NUMBER: 31,990  
 ; REFERENCE/DOCKET NUMBER: 14137-60-1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 206-467-9600  
 ; TELEFAX: 415-576-0300  
 ; INFORMATION FOR SEQ ID NO: 20:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 20 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-465-167A-20

Query Match 69.3%; Score 75.5; DB 1; Length 20;  
 Best Local Similarity 85.7%; Pred. NO. 3.9e-05;  
 Matches 18; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

OY 1 DIEKKICKMEKCSYFNVVNS 21  
 ||||| | | ||||| |  
 Db 1 DIEKKIKMK-KASSYFNVVNS 20

Search completed: January 29, 2002, 10:24:03  
 Job time: 509 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 10:26:36 ; Search time 144.96 Seconds

(without alignments)  
11.035 Million cell updates/sec

Title: US-09-763-397a-7

Perfect score: 109

Sequence: 1 DIEKICKMEKCSVFNVNS 21

Scoring table: BIOSUM62

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : PIR-68:\*

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	109	100.0	405	2 S05428	circumsporozoite p
2	109	100.0	412	1 OZQOAF	circumsporozoite p
3	109	100.0	424	2 A54533	circumsporozoite p
4	109	100.0	442	2 A54529	circumsporozoite p
5	107	98.2	388	2 A39756	circumsporozoite p
6	83	76.1	264	2 A44969	circumsporozoite p
7	83	76.1	332	1 OZQOMB	circumsporozoite p
8	83	76.1	348	1 OZQOMB	circumsporozoite p
9	83	76.1	367	1 OZQOMB	circumsporozoite p
10	77	70.6	429	2 A54504	circumsporozoite p
11	77	70.6	485	2 A60610	circumsporozoite p
12	71	65.1	378	1 OZQOAB	circumsporozoite p
13	71	65.1	378	1 OZQOAB	circumsporozoite p
14	71	65.1	398	1 OZQOAS	circumsporozoite p
15	71	65.1	401	1 OZQOAC	circumsporozoite p
16	71	65.1	419	1 OZQOAM	circumsporozoite p
17	70.5	64.7	388	2 JG6164	circumsporozoite p
18	69	63.3	343	2 A29319	circumsporozoite p
19	69	63.3	367	2 A32068	circumsporozoite p
20	69	63.3	378	1 OZQOAV	circumsporozoite p
21	69	63.3	386	2 A48571	circumsporozoite p
22	69	63.3	387	2 D41156	circumsporozoite p
23	69	63.3	387	2 C41156	circumsporozoite p
24	69	63.3	395	2 A41156	circumsporozoite p
25	65	59.6	351	1 OZQOKU	circumsporozoite p
26	65	59.6	363	1 OZQOKU	circumsporozoite p
27	51	46.8	636	2 A86248	protein T23J18.9 l
28	51	46.8	926	2 T15683	hypothetical prote
29	46	42.2	191	2 S70284	outer surface prot

30	46	42.2	480	2 T20168	hypothetical prote
31	46	42.2	1282	2 J80120	glycoprotein A - m
32	46	42.2	3869	2 A48205	All-1 protein +GRE
33	46	42.2	3968	2 A44265	trichoxin homolog
34	45	41.3	274	2 S59785	probable membrane
35	45	41.3	298	2 A70238	hypothetical prote
36	44	40.4	327	1 RDVZAS	ribonucleoside-dip
37	44	40.4	379	2 B64473	hypothetical prote
38	44	40.4	715	2 JC2222	major surface glyc
39	44	40.4	763	2 T22843	hypothetical prote
40	44	40.4	803	2 S26823	zinc finger protei
41	44	40.4	957	2 T15976	hypothetical prote
42	44	40.4	2339	2 A45597	DNA-directed RNA p
43	43.5	39.9	280	2 S47815	hypothetical 31.8k
44	43.5	39.9	280	2 F86034	hypothetical prote
45	43	39.4	117	2 T07645	pEARL1 protein h

## ALIGNMENTS

RESULT 1  
S05428  
circumsporozoite protein - malaria parasite (Plasmodium falciparum) (isolate NF54)  
C:Species: Plasmodium falciparum  
C:Date: 07-Jun-1990 #sequence: revision 07-Jun-1990 #text\_change 09-Jun-2000  
C:Accession: S05428; A45527; I60657  
R:Campbell, J.R.  
Nucleic Acids Res. 17, 5854, 1989  
A:Title: DNA sequence of the gene encoding a Plasmodium falciparum malaria candidate  
A:Reference number: S05428; MUID:89345189  
A:Accession: S05428  
A>Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-405 <CAM>  
A:Cross-references: EMBL:X15363  
R:Caspers, P.; Gentz, R.; Matile, H.; Pink, J.R.; Sinigaglia, F.  
Mol. Biochem. Parasitol. 37, 275-280, 1989  
A:Title: The circumsporozoite protein gene from NF54, a Plasmodium falciparum isolate  
A:Reference number: A45527; MUID:89364958  
A:Accession: A45527  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-405 <CAS>  
A:Cross-references: GB:M22982; GB:J04650; NID:9160168; PIDN:AAA29527.1; PID:9160169  
R:Lockyer, M.J.; Marsh, K.; Newbold, C.L.  
Mol. Biochem. Parasitol. 37, 275-280, 1989  
A:Title: Wild isolates of Plasmodium falciparum show extensive polymorphism in T cell  
A:Reference number: A60657; MUID:90114334  
A:Accession: I60657  
A>Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 319-336, 354-373 <LOC>  
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology  
F:J29-383/Domain: thrombospondin type 1 repeat homology <THRL>

Query Match 100.0%; Score 109; DB 2; Length 405;  
Best Local Similarity 100.0%; Pred. No. 1.5e-08;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DIEKICKMEKCSVFNVNS 21  
DB 371 DIEKICKMEKCSVFNVNS 391

RESULT 2  
OZQOAF  
circumsporozoite protein - malaria parasite (Plasmodium falciparum) (isolate IMTM22)  
C:Species: Plasmodium falciparum  
C:Date: 15-Nov-1984 #sequence: revision 15-Nov-1984 #text\_change 09-Jun-2000  
C:Accession: A03388  
R:Name, J.B.; Williams, J.L.; McCutchan, T.F.; Weber, J.L.; Wirtz, R.A.; Hochmeyer, W

Science 225, 593-599, 1984  
 A:Title: Structure of the gene encoding the immunodominant surface antigen on the sporozoite  
 A:Reference number: A03388; MUID:84250215  
 A:Accession: A03388  
 A:Molecule type: DNA  
 A:Residues: 1-412 <DM>  
 A:Cross-references: GB:K02194; NID:9160160; PIDN:AAA29524.1; PID:9160161  
 A:Experimental source: clone 7c8  
 C:Comment: Residues 1-16 are the probable signal sequence.  
 C:Comment: There are 41 copies of a 4-residue repeating unit in the middle domain of the protein.  
 C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology <THRI>  
 F:336-390/Domain: thrombospondin type 1 repeat homology

Query Match 100.0%; Score 109; DB 1; Length 412;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-08;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DIEKICKMEKCSVFNNVNS 21  
 |||||  
 Db 378 DIEKICKMEKCSVFNNVNS 398

## RESULT 3

A54533  
 circumsporozoite protein - malaria parasite (Plasmodium falciparum) (strain T4, Thailand)  
 C:Species: Plasmodium falciparum  
 C:Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 09-Jun-2000  
 C:Accession: A54533  
 R:Del Portillo, H.A.; Nussenzweig, R.S.; Enea, V.  
 Mol. Biochem. Parasitol. 24, 289-294, 1987  
 A:Title: Circumsporozoite gene of a Plasmodium falciparum strain from Thailand.  
 A:Reference number: A54533; MUID:87315205  
 A:Accession: A54533  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-424 <DEL>  
 A:Cross-references: GB:M19752; NID:9160216; PIDN:AAA29555.1; PID:9160217  
 C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology  
 F:348-402/Domain: thrombospondin type 1 repeat homology <THRI>

Query Match 100.0%; Score 109; DB 2; Length 424;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-08;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DIEKICKMEKCSVFNNVNS 21  
 |||||  
 Db 390 DIEKICKMEKCSVFNNVNS 410

## RESULT 4

A54529  
 circumsporozoite protein - malaria parasite (Plasmodium falciparum) (strain Wellcome)  
 C:Species: Plasmodium falciparum  
 C:Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 09-Jun-2000  
 C:Accession: A54529  
 R:Lockyer, M.J.; Schwarz, R.T.  
 Mol. Biochem. Parasitol. 22, 101-108, 1987  
 A:Title: Strain variation in the circumsporozoite protein gene of Plasmodium falciparum.  
 A:Reference number: A54529; MUID:87115616  
 A:Accession: A54529  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 1-442 <LOC>  
 A:Cross-references: GB:M15505; NID:9160214; PIDN:AAA29554.1; PID:9160215  
 C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology  
 C:Keywords: tandem repeat  
 F:366-420/Domain: thrombospondin type 1 repeat homology <THRI>

Query Match 100.0%; Score 109; DB 2; Length 442;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-08;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 DIEKICKMEKCSVFNNVNS 21  
 |||||  
 Db 408 DIEKICKMEKCSVFNNVNS 428

## RESULT 5

A39756  
 circumsporozoite protein - Plasmodium reichenowi  
 C:Species: Plasmodium reichenowi  
 C:Date: 14-Feb-1992 #sequence\_revision 14-Feb-1992 #text\_change 20-Aug-1999  
 C:Accession: A39756  
 R:Lal, A.A.; Goldman, I.F.  
 J. Biol. Chem. 266, 6686-6689, 1991  
 A:Title: Circumsporozoite protein gene from Plasmodium reichenowi, a chimpanzee malar  
 A:Reference number: A39756; MUID:91201303  
 A:Accession: A39756  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-388 <LAL>  
 A:Cross-references: GB:M60972; NID:9160228; PIDN:AAA29561.1; PID:9160229  
 C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology  
 F:312-366/Domain: thrombospondin type 1 repeat homology <THRI>

Query Match 98.2%; Score 107; DB 2; Length 388;  
 Best Local Similarity 95.2%; Pred. No. 2.8e-08;  
 Matches 20; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 DIEKICKMEKCSVFNNVNS 21  
 |||||  
 Db 354 DIEKICKMEKCSVFNNVNS 374

## RESULT 6

A44969  
 circumsporozoite protein precursor - Plasmodium yoelii nigeriensis  
 C:Species: Plasmodium yoelii nigeriensis  
 C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 01-Dec-1995  
 C:Accession: A44969  
 R:Colomer-Gould, V.; Enea, V.  
 Mol. Biochem. Parasitol. 43, 51-58, 1990  
 A:Title: Plasmodium yoelii nigeriensis circumsporozoite gene structure and its implic  
 A:Reference number: A44969; MUID:91148645  
 A:Accession: A44969  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-264 <COL>  
 A:Cross-references: GB:M32350  
 C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology  
 F:190-242/Domain: thrombospondin type 1 repeat homology <THRI>

Query Match 76.1%; Score 83; DB 2; Length 264;  
 Best Local Similarity 61.9%; Pred. No. 6.5e-05;  
 Matches 13; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

OY 1 DIEKICKMEKCSVFNNVNS 21  
 ||:||||:||||:|  
 Db 230 DIDTEICKMKDCSIFNIVNS 250

## RESULT 7

OZQMB  
 circumsporozoite protein precursor - Plasmodium berghei (strain NK65)  
 N:Alternate names: sporozoite surface antigen  
 C:Species: Plasmodium berghei  
 C:Date: 30-Sep-1987 #sequence\_revision 28-Jul-1995 #text\_change 16-Jul-1999  
 C:Accession: A44948; A25083; S13446  
 R:Lanar, D.E.  
 Mol. Biochem. Parasitol. 39, 151-154, 1990  
 A:Title: Sequence of the circumsporozoite gene of Plasmodium berghei ANKA clone and N

A:Reference number: A44948; MUID:90158693  
A:Accession: A44948  
A:Molecule type: DNA  
A:Residues: 1-332 <LAN>  
A:Cross-references: GB:M28887  
R:Reichinger, D.J.; Arnott, D.E.; Tam, J.P.; Nussenzweig, V.; Enea, V.  
Mol. Cell. Biol. 6, 3965-3972, 1986  
A:Title: Circumsporozoite protein of Plasmodium berghei: gene cloning and identification  
A:Reference number: A25083; MUID:87089740  
A:Accession: A25083  
A:Molecule type: DNA  
A:Residues: 1-26, 'I', 28-68, 'P', 75-126, 'P', 128-134, 'PPNANDP', 135-332 <ETC>  
A:Cross-references: GB:M1435; MUID:9160245; PIDN:AAA29577.1; PID:9160246  
R:Weber, J.L.; Egan, J.E.; Lyon, J.A.; Wirtz, R.A.; Charoenvit, Y.; Maloy, W.L.; Hockmeyer, J.  
Exp. Parasitol. 63, 295-300, 1987  
A:Title: Plasmodium berghei: cloning of the circumsporozoite protein gene.  
A:Reference number: S13446; MUID:87218962  
A:Accession: S13446  
A:Molecule type: preliminary  
A:Status: preliminary  
A:Residues: 61-122, 'A', 124-332 <MEB>  
A:Cross-references: GB:M25445; MUID:9160177; PIDN:AAA29531.1; PID:9160178  
C:Comment: There are three distinct regions in the mature circumsporozoite protein, the obic membrane-anchoring sequence.  
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology  
C:Keywords: sporozoite; surface antigen; tandem repeat  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:124-332/Product: circumsporozoite protein #status predicted <MAT>  
F:94-189/Region: 8-residue repeats  
F:199-230/Region: 2-residue repeats  
F:258-310/Domain: thrombospondin type 1 repeat homology <THRI>  
  
Query Match 76.1%; Score 83; DB 1; Length 332;  
Best Local Similarity 61.9%; Pred. No. 8e-05;  
Matches 13; Conservative 7; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 DIEKICKMEKCSSEVFNVS 21  
Db 298 DIDREICKMDKCSSEFNVS 318  
||:||||:||||:||||:||||:  
  
RESULT 8  
OZZOBK  
circumsporozoite protein precursor - Plasmodium berghei (strain ANKA clone 2.34L)  
N:Alternate names: sporozoite surface antigen  
C:Species: Plasmodium berghei  
C:Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 16-Jul-1999  
C:Accession: S07873; S12571  
R:Lockyer, M.J.; Davies, C.S.; Subrtler, A.; Sindén, R.E.  
Nucleic Acids Res. 18, 376, 1990  
A:Title: Nucleotide sequence of the Plasmodium berghei circumsporozoite protein gene  
A:Reference number: S07873; MUID:90221834  
A:Accession: S07873  
A:Molecule type: DNA  
A:Residues: 1-348 <LOC>  
A:Cross-references: EMBL:X17606  
R:Lockyer, M.J.  
submitted to the EMBL Data Library, November 1989  
A:Reference number: S12571  
A:Accession: S12571  
A:Molecule type: DNA  
A:Residues: 1-59, 'I', 61-81, 83-348 <LOC2>  
A:Cross-references: EMBL:X17606; MUID:99784; PIDN:CAA35608.1; PID:99785  
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology  
C:Keywords: tandem repeat  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-348/Product: circumsporozoite protein #status predicted <MAT>  
F:94-205/Region: 8-residue repeats  
F:215-247/Region: 2-residue repeats  
F:274-326/Domain: thrombospondin type 1 repeat homology <THRI>

Query Match 76.1%; Score 83; DB 1; Length 348;  
Best Local Similarity 61.9%; Pred. No. 8.3e-05;  
Matches 13; Conservative 7; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 DIEKICKMEKCSSEVFNVS 21  
Db 314 DIDREICKMDKCSSEFNVS 334  
||:||||:||||:||||:||||:  
  
RESULT 9  
OZZOMY  
circumsporozoite protein precursor - Plasmodium yoelii  
N:Alternate names: sporozoite surface antigen  
C:Species: Plasmodium yoelii  
C:Date: 30-Sep-1987 #sequence\_revision 30-Sep-1987 #text\_change 16-Jul-1999  
C:Accession: A26271  
R:Leal, A.A.; de la Cruz, V.F.; Welsh, J.A.; Charoenvit, Y.; Maloy, W.L.; McCutchan, T.  
J. Biol. Chem. 262, 2937-2940, 1987  
A:Title: Structure of the gene encoding the circumsporozoite protein of Plasmodium yoelii  
A:Reference number: A26271; MUID:87137555  
A:Accession: A26271  
A:Molecule type: DNA  
A:Residues: 1-367 <LAL>  
A:Cross-references: GB:J02695; MUID:9160222; PIDN:AAA29558.1; PID:9160223  
C:Comment: There are three distinct regions in the mature circumsporozoite protein, the obic membrane-anchoring sequence.  
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology  
C:Keywords: sporozoite; surface antigen; tandem repeat  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-367/Product: circumsporozoite protein #status predicted <MAT>  
F:139-228/Region: 6-residue repeats (Q-G-P-G-A-P)  
F:229-260/Region: 4-residue repeats (Q-Q-P-P)  
F:293-345/Domain: thrombospondin type 1 repeat homology <THRI>  
  
Query Match 76.1%; Score 83; DB 1; Length 367;  
Best Local Similarity 61.9%; Pred. No. 8.7e-05;  
Matches 13; Conservative 7; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 DIEKICKMEKCSSEVFNVS 21  
Db 333 DIDREICKMDKCSSEFNVS 353  
||:||||:||||:||||:||||:  
  
RESULT 10  
A54504  
circumsporozoite protein - Plasmodium malariae (strain Uganda 1-CDC)  
C:Species: Plasmodium malariae  
C:Date: 06-Oct-1994 #sequence\_revision 04-Nov-1994 #text\_change 20-Aug-1999  
C:Accession: A54504  
R:Leal, A.A.; de la Cruz, V.F.; Campbell, G.H.; Procell, P.M.; Collins, W.E.; McCutchan, T.  
Mol. Biochem. Parasitol. 30, 291-294, 1988  
A:Title: Structure of the circumsporozoite gene of Plasmodium malariae.  
A:Reference number: A54504; MUID:89040027  
A:Accession: A54504  
A:Molecule type: DNA  
A:Residues: 1-429 <LAL>  
A:Cross-references: GB:J03992; MUID:9160220; PIDN:AAA29557.1; PID:9160221  
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology  
C:Keywords: tandem repeat  
F:354-407/Domain: thrombospondin type 1 repeat homology <THRI>  
  
Query Match 70.6%; Score 77; DB 2; Length 429;  
Best Local Similarity 57.1%; Pred. No. 0.00076;  
Matches 12; Conservative 7; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 DIEKICKMEKCSSEVFNVS 21  
Db 395 DLEFICSLDKCSSEFNVS 415  
||:||||:||||:||||:||||:

```
RESULT 11
A60610
Circumsporozoite protein precursor - Plasmodium brasiliannum
C:Species: Plasmodium brasiliannum
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 20-Aug-1999
C:Accession: A60610; A28615
R:di Giovanni, L.; Cochrane, A.H.; Enea, V.
Exp. Parasitol. 70, 373-381, 1990
A:Title: On the evolutionary history of the circumsporozoite protein in plasmodia.
A:Reference number: A60610; MUID:90214818
A:Accession: A60610
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-485 <DIA>
A:Experimental source: Strain Columbian
R:ial, A.A.; de la Cruz, V.F.; Collins, W.E.; Campbell, G.H.; Procell, P.M.; McCutchan,
J. Biol. Chem. 263, 5495-5498, 1988
A:Title: Circumsporozoite protein gene from Plasmodium brasiliannum. Animal reservoirs for
A:Reference number: A28615; MUID:88186854
A:Accession: A28615
A:Molecule type: DNA
A:Residues: 93-485 <IAL>
A:Cross-references: GB:J03203; NID:G160212; PIDN:AA29553.1; PID:G160213
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology
C:Keywords: sporozoite; surface antigen; tandem repeat
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-485/Product: circumsporozoite protein #status predicted <MAT>
F:114-369/Region: 4-residue repeats
F:410-463/Domain: thrombospondin type 1 repeat homology <THR1>

Query Match 70.6%; Score 77; DB 2; Length 485;
Best Local Similarity 57.1%; Pred. No. 0.00085;
Matches 12; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

OY 1 DIEKICKMEKCSYFNVNVS 21
DB 451 DLETEVCTMDKAGIFNVNVS 471

RESULT 12
O2Z0AB
Circumsporozoite protein precursor - Plasmodium cynomolgi (strain Berok)
N:Alternate names: major sporozoite surface antigen
C:Species: Plasmodium cynomolgi
C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 07-Nov-1997
C:Accession: D26255
R:Galinski, M.R.; Arnott, D.E.; Cochrane, A.H.; Barnwell, J.W.; Nussenzweig, R.S.; Enea,
Cell 48, 311-319, 1987
A:Title: The circumsporozoite gene of the Plasmodium cynomolgi complex.
A:Reference number: A90889; MUID:87102878
A:Accession: D26255
A:Molecule type: DNA
A:Residues: 1-378 <GAL>
C:Comment: There are three distinct regions in the mature circumsporozoite protein, the
obscure membrane-anchoring sequence.
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology
C:Keywords: sporozoite; surface antigen; tandem repeat
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-378/Product: circumsporozoite protein #status predicted <MAT>
F:97-192/Region: 9-residue repeats
F:193-268/Region: 16-residue repeats
F:303-356/Domain: thrombospondin type 1 repeat homology <THR1>

Query Match 65.1%; Score 71; DB 1; Length 378;
Best Local Similarity 47.6%; Pred. No. 0.0051;
Matches 10; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

OY 1 DIEKICKMEKCSYFNVNVS 21
DB 344 DLETEVCTMDKAGIFNVNVS 364
```

```
RESULT 13
O2Z0AB
Circumsporozoite protein precursor - Plasmodium cynomolgi (strain London)
N:Alternate names: major sporozoite surface antigen
C:Species: Plasmodium cynomolgi
C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 07-Nov-1997
C:Accession: A26255
R:Galinski, M.R.; Arnott, D.E.; Cochrane, A.H.; Barnwell, J.W.; Nussenzweig, R.S.; Ene
Cell 48, 311-319, 1987
A:Title: The circumsporozoite gene of the Plasmodium cynomolgi complex.
A:Reference number: A90889; MUID:87102878
A:Accession: A26255
A:Molecule type: DNA
A:Residues: 1-378 <GAL>
C:Comment: There are three distinct regions in the mature circumsporozoite protein, t
obscure membrane-anchoring sequence.
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology
C:Keywords: sporozoite; surface antigen; tandem repeat
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-378/Product: circumsporozoite protein #status predicted <MAT>
F:98-211/Region: 6-residue repeats
F:212-277/Region: 11-residue repeats
F:303-356/Domain: thrombospondin type 1 repeat homology <THR1>

Query Match 65.1%; Score 71; DB 1; Length 378;
Best Local Similarity 47.6%; Pred. No. 0.0051;
Matches 10; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

OY 1 DIEKICKMEKCSYFNVNVS 21
DB 344 DLETEVCTMDKAGIFNVNVS 364
```

```
RESULT 14
O2Z0AS
Circumsporozoite protein precursor - Plasmodium cynomolgi (strain Ceylon)
N:Alternate names: major sporozoite surface antigen
C:Species: Plasmodium cynomolgi
C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 07-Nov-1997
C:Accession: C26255
R:Galinski, M.R.; Arnott, D.E.; Cochrane, A.H.; Barnwell, J.W.; Nussenzweig, R.S.; Ene
Cell 48, 311-319, 1987
A:Title: The circumsporozoite gene of the Plasmodium cynomolgi complex.
A:Reference number: A90889; MUID:87102878
A:Accession: C26255
A:Molecule type: DNA
A:Residues: 1-398 <GAL>
C:Comment: There are three distinct regions in the mature circumsporozoite protein, t
obscure membrane-anchoring sequence.
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology
C:Keywords: sporozoite; surface antigen; tandem repeat
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-398/Product: circumsporozoite protein #status predicted <MAT>
F:97-240/Region: 9-residue repeats
F:241-291/Region: 17-residue repeats
F:323-376/Domain: thrombospondin type 1 repeat homology <THR1>

Query Match 65.1%; Score 71; DB 1; Length 398;
Best Local Similarity 47.6%; Pred. No. 0.0054;
Matches 10; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

OY 1 DIEKICKMEKCSYFNVNVS 21
DB 364 DLETEVCTMDKAGIFNVNVS 384

RESULT 15
```



O22QAC  
circumsporozoite protein precursor - Plasmodium cynomolgi (strain Gombak)  
N:Alternate names: major sporozoite surface antigen  
C:Species: Plasmodium cynomolgi  
C:Date: 30-Sep-1987 #sequence\_revision 30-Sep-1987 #text\_change 07-Nov-1997  
E:Accession: E26255  
R:Galinski, M.R.; Arnot, D.E.; Cochran, A.H.; Barnwell, J.W.; Nussenzweig, R.S.; Enea,  
Cell 48, 311-319, 1987  
A:Title: The circumsporozoite gene of the Plasmodium cynomolgi complex.  
A:Reference number: A90889; MUID:87102878  
A:Accession: E26255  
A:Molecule type: DNA  
A:Residues: 1-401 <GAL>  
C:Comment: There are three distinct regions in the mature circumsporozoite protein, the  
obscure membrane-anchoring sequence.  
C:Comment: There are 17 tandem copies of the 11-residue repeat D/G-A-A-A-G-G-G-N.  
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology  
C:Keywords: sporozoite; surface antigen; tandem repeat  
F:1-19/Domains: signal sequence #status predicted <SIG>  
F:20-401/Product: circumsporozoite protein #status predicted <MAT>  
F:98-278/Region: 11-residue repeats  
F:326-379/Domains: thrombospondin type 1 repeat homology <THR1>

	Query Match	65.1%	Score 71:	DB 1:	Length 401:
	Best Local Similarity	47.6%	Pred. No.	0.0054:	
	Matches 10:	Conservative 8:	Mismatches 3:	Indels 0:	Gaps 0:
Qy	1 DIEKRICKEKCCSVFENYVNS	21			
	I I : : I : I I I : I I I I :				
Dd	367 DLEVECTMDCKAGLIFNVNS	387			

Search completed: January 29, 2002, 10:26:36  
Job time: 647 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 11:13:47 ; Search time 80.65 Seconds

(without alignments)  
8.638 Million cell updates/sec

Title: US-09-763-397A-20

Perfect score: 96

Sequence: 1 NEREDERTLTKEYEDIVLK 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SWISSProt\_39:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	96	100.0	1435	1	EBAL_PLAFC
2	49	51.0	442	1	BAM_DROME
3	49	51.0	598	1	YDM3_SCHPO
4	46	47.9	213	1	VD09_VACCC
5	46	47.9	213	1	VD09_VACCV
6	46	47.9	213	1	VD09_VACV
7	46	47.9	213	1	VD09_VARY
8	45	46.9	513	1	TBA1_NEUCR
9	45	46.9	1282	1	GAL7_BACSU
10	44	45.8	332	1	TP2M_DICDI
11	44	45.8	451	1	LDHD_LACDE
12	44	45.8	985	1	TBA2_HOMAM
13	43.5	44.8	731	1	BAF1_YEAST
14	43.5	44.8	943	1	DIP2_YEAST
15	42	43.8	208	1	RL4_MYCA
16	42	43.8	239	1	FRPB_WOLSU
17	42	43.8	274	1	Y059_CAEEL
18	42	43.8	603	1	HEM1_CHICK
19	41	42.7	101	1	Y701_MYCV
20	41	42.7	196	1	CAIE_ECOLI
21	41	42.7	450	1	TBA_NOTVI
22	41	42.7	463	1	CMGA_MOUSE
23	41	42.7	466	1	CMGA_RAT
24	41	42.7	668	1	UJ52_HCVNA
25	41	42.7	715	1	SMC7_SCHMA
26	41	42.7	776	1	YOFB_BACSU
27	41	42.7	776	1	TOP1_RICPR
28	41	42.7	1770	1	RI15_YEAST
29	40.5	42.2	457	1	CMGA_HUMAN
30	40	41.7	240	1	TBA_OCTVU
31	40	41.7	347	1	YCEG_HAEIN
32	40	41.7	353	1	BMP2_CHICK
33	40	41.7	393	1	BMP2_RAT

34	40	41.7	394	1	BMP2_MOUSE
35	40	41.7	395	1	BMP2_RABIT
36	40	41.7	396	1	BMP2_DAMDA
37	40	41.7	396	1	BMP2_HUMAN
38	40	41.7	398	1	BMPA_XENLA
39	40	41.7	398	1	BMPB_XENLA
40	40	41.7	412	1	TBA1_CHICK
41	40	41.7	444	1	TBA_ONCKE
42	40	41.7	448	1	TBA4_HUMAN
43	40	41.7	448	1	TBA5_CHICK
44	40	41.7	449	1	CMGA_BOVIN
45	40	41.7	449	1	TBA6_MOUSE

## ALIGNMENTS

RESULT 1	EBAL_PLAFC	STANDARD:	PRT: 1435 AA.
ID	EBAL_PLAFC		
AC	P19214:		
DT	01-NOV-1990 (Rel. 16, Created)		
DT	01-FEB-1996 (Rel. 33, Last sequence update)		
DT	01-FEB-1996 (Rel. 33, Last annotation update)		
DE	ERYTHROCYTE-BINDING ANTIGEN EBA-175.		
OS	Plasmodium falciparum (isolate Camp / Malaysia).		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_TaxID=5835;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=90377299; PubMed=2204835;		
RA	Sim B.K.L.:		
RT	"Sequence conservation of a functional domain of erythrocyte binding		
RT	antigen I/5 in Plasmodium falciparum.";		
RL	Mol. Biochem. Parasitol. 41:293-296(1990).		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sdb.ch/announce/">http://www.isb-sdb.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sdb.ch">license@isb-sdb.ch</a> ).		
CC	-----		
DR	EMBL: X52524; CAA36756.1; -.		
DR	PIR: S11561; S11561.		
KW	Antigen.		
FT	DOMAIN 159 1104		
FT	ESSENTIAL FOR BINDING TO		
FT	ERYTHROCYTES.		
FT	VARIANT 1031 1031 E -> V (IN STRAINS FCR-3 AND ITG).		
FT	SEQUENCE 1435 AA; 167389 MW; 32A4309021BIC3D6 CRC64;		
SO	SEQUENCE		
Query Match	100.0%; Score 96; DB 1; Length 1435;		
Best Local Similarity	100.0%; Pred. No. 1.5e+06;		
Matches	19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Oy	1 NEREDERTLTKEYEDIVLK 19		
Db	1069 NEREDERTLTKEYEDIVLK 1087		
RESULT 2	BAM_DROME		
ID	BAM_DROME		
AC	P22745: O9BVO0:		
DT	01-AUG-1991 (Rel. 19, Created)		
DT	20-AUG-2001 (Rel. 40, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	BAG-OF-MARPLES PROTEIN (BAM PROTEIN).		
GN	BAM OR CG10422.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		

CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC STRAIN=FS(3)NEO61; TISSUE=ovary;  
 CC MEDLINE=91122627; PubMed=2279698;  
 RA McKeatin D.M., Spradling A.C.;  
 RT "Bag-of-marbles: a Drosophila gene required to initiate both male and  
 female gametogenesis.";  
 RL Genes Dev. 4:2242-2251(1990).  
 RN [2]  
 RP REVISIONS TO 99-113.  
 RA McKeatin D.M.;  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 CC MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanalides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.-H., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer V.G., Champe M., Pfeiffer B.D.,  
 RA Abil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Berooz P.V., Bernier B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Flockerzi A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclob J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-T., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 CC -1- FUNCTION: REQUIRED TO INITIATE BOTH MALE AND FEMALE GAMETOGENESIS.  
 CC MAY REGULATE CYSTOBLAST CELL DIVISIONS.  
 CC -1- TISSUE SPECIFICITY: IN CYSTOBLASTS AND/OR VERY EARLY CYSTOCYTES IN  
 CC OVARY AND TESTIS.  
 CC -1- SIMILARITY: LOW, WITH THE DROSOPHILA OVARIAN TUMOR GENE PROTEIN.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X56202; CAA39662.1; -  
 DR EMBL: AF003751; AAF56427.1; -

DR PIR: A36469; A36469.  
 DR FlyBase; FBgn0000158; bam.  
 FT DOMAIN 402 434 GLU/SER/PRO/THR-RICH (EST REGION).  
 FT CONFLICT 239 239 A -> S (IN REF. 1).  
 FT CONFLICT 439 439 R -> A (IN REF. 1).  
 SO SEQUENCE 442 AA; 50275 MW; DFE2BFCFC249FE254 CRC64;  
 Query Match 51.0%; Score 49; DB 1; Length 442;  
 Best Local Similarity 66.7%; Pred. No. 4.4;  
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 NERDEPRTLKEYED 15  
 DB 35 NENEDPRKATCEYED 49  
 RESULT 3  
 YDWS\_SCHPO STANDARD; PRT; 598 AA.  
 AC 013910;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE HYPOTHETICAL 67.7 KDA PROTEIN C23C11.03 IN CHROMOSOME I.  
 GN SPAC23C11.03.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Brown D., Churcher C.M., Barrell B.G., Rajandream M.A., Wood V.,  
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE MPPID FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z98559; CAB1156.1; -  
 KW Hypothetical protein.  
 FT DOMAIN 237 240 POLY-LYS.  
 FT DOMAIN 462 465 POLY-PRO.  
 FT DOMAIN 532 538 POLY-ARG.  
 SO SEQUENCE 598 AA; 67707 MW; 36439BA7106FEF99 CRC64;  
 Query Match 51.0%; Score 49; DB 1; Length 598;  
 Best Local Similarity 47.4%; Pred. No. 6.2;  
 Matches 9; Conservative 5; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 NERDEPRTLKEYEDYLVK 19  
 DB 401 NENKSQSLAEYEEERFLK 419  
 RESULT 4  
 VD09\_VACCC STANDARD; PRT; 213 AA.  
 AC P21011;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PROTEIN D9.  
 GN DBR.  
 OS Vaccinia virus (strain Copenhagen).

```

OC  Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC  Orthopoxvirus.
OX  NCBI_TaxID=10249;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=91021027; PubMed=2219722;
RA  Goebel S.J., Johnson G.P., Perkins M.E., Davis S.W., Winslow J.P.,
RA  Paoletti E.;
RT  "The complete DNA sequence of vaccinia virus.";
RL  Virology 179:247-266(1990).
RN  [2]
RP  COMPLETE GENOME.
RA  Goebel S.J., Johnson G.P., Perkins M.E., Davis S.W., Winslow J.P.,
RA  Paoletti E.;
RT  "Appendix to 'The complete DNA sequence of vaccinia virus.'";
RL  Virology 179:517-563(1990).
CC  -1- SIMILARITY: BELONGS TO THE NUDIX HYDROLASE FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: M35027; AAA48108.1; -
DR  PIR: H42515; H42515.
DR  InterPro: IPR000086; NUDIX_hydrolase.
DR  InterPro: IPR003300; Viral_VD9.
DR  Pfam: PF00293; mult; 1.
DR  PRINTS: PR00502; MUTDOMAIN.
DR  PROSITE: PS00893; NUDIX; 1.
KW  Hydrolase.
FT  DOMAIN 111 132 NUDIX BOX.
FT  SEQUENCE 213 AA; 25024 MW; 747688883C43EF7E CRC64;

Query Match 47.9%; Score 46; DB 1: Length 213;
Best Local Similarity 44.4%; Pred. No. 5.6;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 2 EREDERTLTKEDYIVLK 19
Db 129 EESDERITVKEFGNVILK 146
1 111 11: :::11

RESULT 5
VD09_VACCV STANDARD; PRT; 213 AA.
ID VD09_VACCV
AC P04311;
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN D9.
GN D9R.
OS Vaccinia virus (strain WR).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10254;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86291159; PubMed=3739227;
RA Niles E.G., Condit R.C., Caro P., Davidson K., Matusick L., Seto J.;
RT "Nucleotide sequence and genetic map of the 16-kb vaccinia virus
RT HindIII D fragment.";
RL Virology 153:96-112(1986).
CC -1- SIMILARITY: BELONGS TO THE NUDIX HYDROLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR  EMBL: M15058; AAA48265.1; -
DR  PIR: A03885; Q0VZ15.
DR  InterPro: IPR000086; NUDIX_hydrolase.
DR  InterPro: IPR003300; Viral_VD9.
DR  Pfam: PF00293; mult; 1.
DR  PRINTS: PR00502; MUTDOMAIN.
DR  PROSITE: PS00893; NUDIX; 1.
KW Hydrolase.
FT  DOMAIN 111 132 NUDIX BOX.
FT  SEQUENCE 213 AA; 24994 MW; 306288883C5899A5 CRC64;

Query Match 47.9%; Score 46; DB 1: Length 213;
Best Local Similarity 44.4%; Pred. No. 5.6;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 2 EREDERTLTKEDYIVLK 19
Db 129 EESDERITVKEFGNVILK 146
1 111 11: :::11

RESULT 6
VD09_VARY STANDARD; PRT; 213 AA.
ID VD09_VARY
AC P33070;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN D9.
GN D9R OR F9R.
OS Variola virus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10255;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=INDIA-1967 / ISOLATE IND3;
RX MEDLINE=93190624; PubMed=8383392;
RA Shchelkunov S.N., Blinov V.M., Tolmenin A.V., Marennikova S.S.,
RA Kolykhalov A.A., Frolov I.V., Chizhikov V.E., Gytarov V.V.,
RA Gashikov P.V., Belanov E.F., Belavin P.A., Resenchuk S.M.,
RA Andzhaparidze O.G., Sandakhchiev L.S.;
RT "Nucleotide sequence analysis of variola virus HindIII M, L, I genome
RT fragments.";
RL Virus Res. 27:25-35(1993).
RN [2]
RP COMPLETE GENOME.
RX STRAIN=INDIA-1967 / ISOLATE IND3;
RX MEDLINE=93202281; PubMed=8384129;
RA Shchelkunov S.N., Blinov V.M., Sandakhchiev L.S.;
RT "Genes of variola and vaccinia viruses necessary to overcome the host
RT protective mechanisms.";
RL FEBS Lett. 319:80-83(1993).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=BANGLADESH-1975;
RX MEDLINE=94088747; PubMed=8264798;
RA Massung R.F., Esposito J.J., Liu L., Qi J., Utterback T.R.,
RA Knight J.C., Aubin L., Yuran T.E., Parsons J.M., Loparev V.N.,
RA Selivanov N.A., Cavallaro K.F., Kerlavage A.R., Mahy B.W.J.,
RA Venter C.J.;
RT "Potential virulence determinants in terminal regions of variola
RT smallpox virus genome.";
RL Nature 366:748-751(1993).
CC -1- SIMILARITY: BELONGS TO THE NUDIX HYDROLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X67119; CAA47598.1; -;  
DR EMBL; X69198; CAA49040.1; -;  
DR EMBL; L22579; AAA60847.1; -;  
DR PIR; F36847; F36847.  
DR PIR; S33113; S33113.  
DR InterPro; IPR000086; NUDIX\_hydrolase.  
DR InterPro; IPR003300; VITAL\_VD9.  
DR Pfam; PF00293; mult; 1.  
DR PRINTS; PR00502; MUTDOMAIN.  
DR PROSITE; PS00893; NUDIX; 1.  
KW Hydrolase.  
FT DOMAIN 111 132 NUDIX BOX.  
SQ SEQUENCE 213 AA; 24981 MW; 32AD6A5AF3B899A6 CRC64;  
  
OY 2 EREDRTLKREYEDIVLK 19  
1 111 11: : : : :  
DB 129 EESDERITKEFGNVILK 146  
  
RESULT 7  
TBAL\_NEUCR STANDARD; PRT; 453 AA.  
AC P38668;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE TUBULIN ALPHA-A CHAIN.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_Taxid=5141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ST. LAWRENCE 74 / SL 74 / ORS 6A;  
RX MEDLINE=97306636; PubMed=9163903;  
RA Monnat J., Ortega Perez R., Turian G.;  
RT "Molecular cloning and expression studies of two divergent  
RT alpha-tubulin genes in Neurospora crassa";  
RL FEMS Microbiol. Lett. 150:33-41(1997).  
CC -1- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES. IT  
CC BINDS TWO MOLES OF GTP, ONE AT AN EXCHANGEABLE SITE ON THE BETA  
CC CHAIN AND ONE AT A NONEXCHANGEABLE SITE ON THE ALPHA-CHAIN.  
CC -1- SUBUNIT: DIMER OF ALPHA AND BETA CHAINS.  
CC -1- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X79403; CAA55940.1; -;  
DR PIR; S45050; S45050.  
DR InterPro; IPR000217; Tubulin.  
DR InterPro; IPR003008; Tubulin\_fts2.  
DR Pfam; PF00091; tubulin; 1.  
DR PRINTS; PR01162; TUBULIN.  
DR PRINTS; PR01162; ALPHATUBULIN.  
DR PROSITE; PS00227; TUBULIN; 1.  
KW Microtubules; GTP-binding; Multigene family.  
FT NP\_BIND 142 148 GTP (POTENTIAL).

FT SITE 453 453 INVOLVED IN POLYMERIZATION.  
SQ SEQUENCE 453 AA; 50140 MW; CAF292C758F8038 CRC64;  
  
OY 3 REDRTLKREYED 16  
1 111 11: : : : :  
DB 427 REDRTLKREYEDV 440  
  
Query Match 47.9%; Score 46; DB 1; Length 453;  
Best Local Similarity 57.1%; Pred. No. 13;  
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
  
RESULT 8  
GAL7\_BACSU STANDARD; PRT; 513 AA.  
ID GAL7\_BACSU  
AC P39575;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE GALACTOSE-1-PHOSPHATE URIDYLTRANSFERASE (EC 2.7.7.10) (GAL-1-P  
DE URIDYLTRANSFERASE).  
GN GALT OR IPA-36D.  
CN GALT OR IPA-36D.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_Taxid=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=95020537; PubMed=7934828;  
RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,  
RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,  
RA Presceon E., Santana M., Schneider E., Schweizer J., Yvetes A.,  
RA Rapoport G., Danchin A.;  
RT "Bacillus subtilis genome project: Cloning and sequencing of the 97  
RT kb region from 325 degrees to 333 degrees";  
RL Mol. Microbiol. 10:371-384(1993).  
CC -1- CATALYTIC ACTIVITY: UTP + ALPHA-D-GALACTOSE 1-PHOSPHATE =  
CC UDP-GALACTOSE + PYROPHOSPHATE.  
CC -1- PATHWAY: GALACTOSE METABOLISM.  
CC -1- SIMILARITY: BELONGS TO THE GALACTOSE-1-PHOSPHATE  
CC URIDYLTRANSFERASE FAMILY 2.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X73124; CAA51592.1; -;  
DR EMBL; Z99123; CAB15645.1; -;  
DR PIR; S39691; S39691.  
DR Subtilist; BG10582; galT.  
DR InterPro; IPR000880; Galp UDP\_transf.  
DR InterPro; IPR000766; Galp UDP\_transf\_II.  
DR Pfam; PF01087; Galp\_UDP\_transf; 1.  
DR PROSITE; PS01163; GAL\_P\_UDP\_TRANSF\_II; 1.  
KW Transferase; Nucleotidyltransferase; Galactose metabolism;  
KW Complete proteome.  
SQ SEQUENCE 513 AA; 58605 MW; EBBF2A57A3987532 CRC64;  
  
OY 5 DERTITKEVEDIVLK 19  
1 111 11: : : : :  
DB 455 EKRTITKEVEDIVLK 469  
  
Query Match 46.9%; Score 45; DB 1; Length 513;  
Best Local Similarity 53.3%; Pred. No. 21;  
Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

```

RESULT 9
ID TP2M_DICDI STANDARD: PRT: 1282 AA.
AC P90520:
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA TOPOISOMERASE II, MITOCHONDRIAL PRECURSOR (EC 5.99.1.3).
GN TOPI OR TOPIA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3:
RX MEDLINE=97320633; PubMed=9177484;
RA Komori K., Kuroe K., Yanagisawa K., Tanaka Y.;
RT "Cloning and characterization of the gene encoding a mitochondrially
RT localized DNA topoisomerase II in Dictyostelium discoideum. Western
RT blot analysis."
RL Blochim. Biophys. Acta 1352:63-72(1997).
CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
CC MAKES DOUBLE-STRAND BREAKS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
CC OF DOUBLE-STRANDED DNA.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC -1- DEVELOPMENTAL STAGE: PRESENT IN GROWTH PHASE AND DURING
CC DEVELOPMENT, ALTHOUGH LEVELS DECLINED AS DEVELOPMENT PROCEEDED.
CC -1- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC
DR EMBL: D82024; BAA11510.1; -.
DR Dictydb: DD05117; topb.
DR InterPro: IPR001241; DNA_topoisolt.
DR InterPro: IPR002305; DNA_topoisolt.
DR InterPro: IPR003594; HATPase_c.
DR Pfam: PF00204; DNA_topoisolt; 2.
DR Pfam: PF00521; DNA_topoisolt; 1.
DR Pfam: PF02518; HATPase_c; 1.
DR PRINTS: PR00418; TP12FAMILY.
DR PRINTS: PR01158; TOPISMASREII.
DR PRODOM: PD000616; DNA_topoisolt; 1.
DR SMART: SM00387; HATPase_c; 1.
DR SMART: SM00433; TOP2c; 1.
DR SMART: SM00434; TOP2c; 1.
DR PROSITE: PS00177; TOPOISOMERASE_II; 1.
KW Isomerase; Topoisomerase; DNA-binding; ATP-binding;
KW Mitochondrion; Transist peptide.
FT TRANSIT 1 35 MITOCHONDRION (POTENTIAL).
FT CHAIN 36 1282 DNA TOPOISOMERASE II.
FT NP_BIND 253 258 ATP (POTENTIAL).
FT ACT_SITE 899 899 DNA CLEAVAGE (BY SIMILARITY).
FT DOMAIN 147 154 POLY-THR.
FT DOMAIN 1252 1256 POLY-SER.
FT DOMAIN 1260 1265 POLY-SER.
SQ SEQUENCE 1282 AA; 145866 MW; 7E21DEBF3D069951 CRC64;

```

```

Query Match 46.9%; Score 45; DB 1; Length 1282;
Best Local Similarity 47.4%; Pred. No. 56;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
Oy 1 NEREDERTLTKEYEDIVLK 19

```

```

DB 75 NEKISEKTTTRKIEDIYOK 93
||: |:| |:| |:| |
RESULT 10
ID LDHD_LACDE STANDARD: PRT: 332 AA.
AC P26297;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE D-LACTATE DEHYDROGENASE (EC 1.1.1.28) (D-LDH).
GN LDHA.
OS Lactobacillus delbrueckii (subsp. bulgaricus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1585;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIB 11778;
RX MEDLINE=92008688; PubMed=1915894;
RA Bernard N., Ferrain T., Garmyn D., Hols P., Delcour J.;
RT "Cloning of the D-lactate dehydrogenase gene from Lactobacillus
RT delbrueckii subsp. bulgaricus by complementation in Escherichia
RT coli."
RL FEBS Lett. 290:61-64(1991).
RN [2]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=92235079; PubMed=1569100;
RA Kochhar S., Hunziker P., Leong-Morgenthaler P.M., Hottinger H.;
RT "Primary structure, physicochemical properties, and chemical
RT modification of NAD(+)-dependent D-lactate dehydrogenase. Evidence
RT for the presence of Arg-235, His-303, Tyr-101, and Trp-19 at or near
RT the active site."
RL J. Biol. Chem. 267:8499-8513(1992).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92304298; PubMed=1610363;
RA Kochhar S., Chuard N., Hottinger H.;
RT "Cloning and overexpression of the Lactobacillus bulgaricus NAD(+)-
RT dependent D-lactate dehydrogenase gene in Escherichia
RT coli: purification and characterization of the recombinant enzyme."
RL Biochem. Biophys. Res. Commun. 185:705-712(1992).
RN [4]
RP SIMILARITY TO OTHER ENZYMES OF THIS FAMILY.
RX MEDLINE=92231983; PubMed=1567457;
RA Kochhar S., Hunziker P., Leong-Morgenthaler P.M., Hottinger H.;
RT "Evolutionary relationship of NAD(+)-dependent D-lactate
RT dehydrogenase: comparison of primary structure of 2-hydroxy acid
RT dehydrogenases."
RL Biochem. Biophys. Res. Commun. 184:60-66(1992).
RN [5]
RP MUTAGENESIS OF HIS-205; ARG-235; ASP-259; GLU-264 AND HIS-296.
RX MEDLINE=97217445; PubMed=9063466;
RA Bernard N., Johnson K., Geipi J.L., Alvarez J.A., Ferrain T.,
RA Garmyn D., Hols P., Cortes A., Clarke A.R., Holbrook J.J., Delcour J.;
RT "D-2-hydroxy-4-methylvalerate dehydrogenase from Lactobacillus
RT delbrueckii subsp. bulgaricus. II. Mutagenic analysis of
RT catalytically important residues."
RL Eur. J. Biochem. 244:213-219(1997).
RN [6]
RP 3D-STRUCTURE MODELING.
RA Vinals C., de Bolle X., Deleureux E., Feytmans E.;
RL Submitted (JAN-1995) to the EMBL/Genbank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: D-LACTATE + NAD(+) = PYRUVATE + NADH.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO THE D-ISOMER SPECIFIC 2-HYDROXYACID
CC DEHYDROGENASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X60220; CAA42781.1; -  
DR EMBL: M65224; AAA25246.1; -  
DR PIR: S17556; S17556.  
DR PIR: JN0245; JN0245.  
DR PIR: A38094; A38094.  
DR PDB: 1DLD; 31-MAR-95.  
DR InterPro: IPR002162; D\_2\_hydroxyacid\_DH.  
DR Pfam: PF00389; 2-Hacid\_DH.1.  
DR PROSITE: PS00065; D\_2-HYDROXYACID\_DH.1; 1.  
DR PROSITE: PS00670; D\_2-HYDROXYACID\_DH.2; 1.  
DR PROSITE: PS00671; D\_2-HYDROXYACID\_DH.3; 1.  
KW Oxidoreductase; NAD; 3D-structure.  
FT INIT MET 0  
FT NP\_BIND 147 175 NAD (BY SIMILARITY).  
FT ACT\_SITE 235 235 SUBSTRATE-BINDING (BY SIMILARITY).  
FT ACT\_SITE 264 264 BY SIMILARITY.  
FT ACT\_SITE 296 296 BY SIMILARITY.  
FT MUTAGEN 205 205 H->Q: INCREASE OF ACTIVITY.  
FT MUTAGEN 235 235 R->K: DECREASE OF ACTIVITY.  
FT MUTAGEN 259 259 D->N: DECREASE OF ACTIVITY.  
FT MUTAGEN 264 264 E->Q: DECREASE OF ACTIVITY.  
FT MUTAGEN 296 296 H->Q: 90% LOSS OF ACTIVITY.  
FT CONFLICT 40 40 A -> V (IN REF. 1).  
FT CONFLICT 116 116 R -> A (IN REF. 1).  
FT CONFLICT 121 121 A -> D (IN REF. 1).  
FT CONFLICT 151 151 V -> I (IN REF. 1).  
FT CONFLICT 173 173 A -> T (IN REF. 1).  
FT CONFLICT 219 219 K -> E (IN REF. 1).  
FT CONFLICT 253 253 V -> I (IN REF. 1).  
FT CONFLICT 267 267 V -> I (IN REF. 1).  
FT CONFLICT 272 272 R -> W (IN REF. 1).  
SQ SEQUENCE 332 AA; 36773 MW; D776CA0746589449 CRC64;

QY 3 REDERTLTKREVED 15  
Db 9 REDEKPLKEMED 21

Query Match 45.8%; Score 44; DB 1; Length 332;  
Best Local Similarity 61.5%; Pred. No. 18;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

RESULT 11  
TBA2\_HOMAM STANDARD; PRT; A51 AA.  
AC 094570;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE TUBULIN ALPHA-2 CHAIN (ALPHA-TI TUBULIN).  
OS Homarus americanus (American lobster).  
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;  
OC Nephropoidea; Nephropidae; Homarus.  
OX NCBI\_Taxid=6706;  
RN RP SEQUENCE FROM N.A.  
RA MEDLINE=96257217; PubMed=8666270;  
RA Demers D.M., Metcalf A.E., Talbot P., Hyman B.C.;  
RT "Multiple lobster tubulin isoforms are encoded by a single gene  
family";  
RL Gene 171;185-191(1996).  
CC -1- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES. IT  
CC BINDS TWO MOLES OF GTP, ONE AT AN EXCHANGEABLE SITE ON THE BETA  
CC CHAIN AND ONE AT A NONEXCHANGEABLE SITE ON THE ALPHA-CHAIN.  
CC -1- SUBUNIT: DIMER OF ALPHA AND BETA CHAINS.  
CC -1- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: U66318; AAB07481.1; -  
DR InterPro: IPR000217; Tubulin.  
DR InterPro: IPR003008; Tubulin\_Ftsz.  
DR Pfam: PF00091; tubulin; 1.  
DR PRINTS: PR01161; TUBULIN.  
DR PRINTS: PR01162; ALPHATUBULIN.  
DR PROSITE: PS00227; TUBULIN; 1.  
KW Microtubules; GTP-binding; Multigene family.  
FT NP\_BIND 142 148 GTP (POTENTIAL).  
FT SITE 451 451 INVOLVED IN POLYMERIZATION.  
SQ SEQUENCE 451 AA; 50250 MW; 5D7751B233035856 CRC64;

QY 3 REDERTLTKREVED 16  
Db 422 REDATLTKREVED 435

Query Match 45.8%; Score 44; DB 1; Length 451;  
Best Local Similarity 57.1%; Pred. No. 25;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

RESULT 12  
NAH1\_YEAST STANDARD; PRT; 985 AA.  
ID NAH1\_YEAST  
AC 099271;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE NA(+)/H(+) ANTI-PORTER.  
GN NAH1 OR YLR138W OR L3149 OR L9606.4.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_Taxid=4932;  
RN RP SEQUENCE FROM N.A.  
RC STRAIN=5288C / AB972;  
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
RA Favell A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,  
RA Hallsworth K., Langston Y., Hillier L., Jier M., Johnson D.,  
RA Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,  
RA Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,  
RA Taich A., Trevaskis E., Vignati D., Wilcox L., Woldman P., Vaudin M.,  
RA Wilson R., Waterston R.,  
RL Submitted (APR-1996) to the EMBL/Genbank/DBJ databases.  
RN RP SEQUENCE FROM N.A.  
RC STRAIN=5288C;  
RA Delius H., Hedling U.;  
RL Submitted (SEP-1995) to the EMBL/Genbank/DBJ databases.  
RN RP SEQUENCE OF 145-985 FROM N.A.  
RA Rieger M., Mueller-Auer S., Brueckner M.;  
RL Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.  
CC -1- FUNCTION: SODIUM EXPORT FROM CELL. TAKES UP EXTERNAL PROTONS IN  
CC EXCHANGE FOR INTERNAL SODIUM IONS (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -1- SIMILARITY: BELONGS TO THE FUNGAL NA(+)/H(+) EXCHANGER FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial





DIP2\_YEAST  
ID DIP2\_YEAST STANDARD: PRT: 943 AA.  
AC Q12220; Q05386;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE DOM34 INTERACTING PROTEIN 2.  
GN DIP2 OR YLR129W OR L9233.1 OR L3116.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP STRAIN-S288C / AB972;  
RC Johnson M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
RA Favell A., Fulton L., Gattung S., Greco T., Kirszen J., Kucaba T.,  
RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,  
RA Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,  
RA Miller N., Nhan M., Pauley A., Peluso D., Rifkin L., Riles L.,  
RA Tach A., Trevasakis E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,  
RA Wilson R., Waterston R.;  
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP STRAIN-S288C;  
RC STRAIN-S288C;  
RA Delius H.;  
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 1-458 FROM N.A.  
RC STRAIN-FY23 / RD005;  
RA Verhasselt P., Voet M., Volckaert G.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: CONTAINS 10 WD REPEATS (TRP-ASP DOMAINS).  
CC -1- SIMILARITY: TO S.POMBE SPEC3D5.12.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U53877; AAB82375.1; -;  
DR EMBL: U53881; AAB82402.1; -;  
DR EMBL: X91258; CAA62640.1; -;  
DR EMBL: Z73301; CAA97699.1; -;  
DR EMBL: Z73302; CAA97700.1; -;  
DR EMBL: X89514; CAA61707.1; -;  
DR SGD: S0004119; DIP2.  
DR InterPro: IPR001680; WD40.  
DR Pfam: PF00400; WD40; 12.  
DR PRINTS: PR00320; GPROTEINRPT.  
DR SMART: SMO0320; WD40; 11.  
DR PROSITE: PS00678; WD\_REPEATS\_1; 4.  
DR PROSITE: PS00082; WD\_REPEATS\_2; 8.  
DR PROSITE: PSS0294; WD\_REPEATS\_REGION; 1.  
KW Repeat; WD repeat.  
FT REPEAT 77 WD 1.  
FT REPEAT 119 WD 2.  
FT REPEAT 161 WD 3.  
FT REPEAT 202 WD 4.  
FT REPEAT 389 WD 5.  
FT REPEAT 428 WD 6.  
FT REPEAT 471 WD 7.  
FT REPEAT 571 WD 8.  
FT REPEAT 613 WD 9.  
FT REPEAT 655 WD 10.  
SQ SEQUENCE 943 AA; 106342 MW; 7D9AECTAF6A9C740 CRC64;

Query Match 44.8%; Score 43; DB 1; Length 943;

Best Local Similarity 47.1%; Pred. No. 80;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
QY 2 EREDERTLTKEVEDIVL 18  
DB 694 EEKEKELEEEQVEDIVL 710  
RESULT 15  
ID RL4\_MYCCA STANDARD: PRT: 208 AA.  
AC P10135;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 01-AUG-1992 (Rel. 23, Last annotation update)  
DE 50S RIBOSOMAL PROTEIN L4.  
GN RP.LD.  
OS Mycoplasma capricolum.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Molluscites;  
OC Entomoplasmataceae.  
OX NCBI\_TaxID=2095;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 27343 / KID;  
RX MEDLINE=88142549; PubMed=3481422;  
RA Okubo S., Muto A., Kawauchi Y., Yamao F., Osawa S.;  
RT "The ribosomal protein gene cluster of Mycoplasma capricolum";  
RL Mol. Gen. Genet. 210:314-322(1987).  
CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY AND SPECIFICALLY TO 23S  
CC RNA (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE L4P FAMILY OF RIBOSOMAL PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X06414; CAA29705.1; -;  
DR PIR: S02832; R5YMAC.  
DR InterPro: IPR002136; Ribosomal\_L4/L1E.  
DR Pfam: PF00573; Ribosomal\_L4; 1.  
KW Ribosomal protein; rRNA-binding.  
SQ SEQUENCE 208 AA; 23109 MW; D6D3ECDA54ICB73A CRC64;

Query Match 43.8%; Score 42; DB 1; Length 208;  
Best Local Similarity 50.0%; Pred. No. 22;  
Matches 9; Conservative 6; Mismatches 1; Indels 2; Gaps 1;  
QY 4 EDERTL--TKEVEDIVK 19  
DB 149 DDQKTLVTKKEELVVK 166

Search completed: January 29, 2002, 11:13:48  
Job time: 824 sec



THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 11:12:14 ; Search time 285.36 Seconds  
(without alignments)  
9.739 Million cell updates/sec

Title: US-09-763-397A-20  
Perfect score: 96  
Sequence: 1 NEREDERTLTKEYEDIVLK 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_17.\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	96	100.0	1433	5	Q9N9G9
2	96	100.0	1475	5	Q25842
3	89	92.7	1421	5	Q05644
4	89	92.7	1435	5	Q9NG63
5	50	52.1	368	10	Q9ZUX9
6	48	50.0	217	2	Q51118
7	48	50.0	669	2	Q9FCS0
8	48	50.0	669	2	Q9FCR5
9	48	50.0	680	2	Q51283
10	48	50.0	685	2	Q9AKU7
11	48	50.0	689	2	Q51288
12	48	50.0	697	2	Q9EXC8
13	48	50.0	702	2	Q9FCR9
14	48	50.0	707	2	Q9EXB0
15	48	50.0	722	2	Q9AKU8
16	47	49.0	517	10	Q9XFJ4
17	46	47.9	213	12	O57212
18	46	47.9	384	10	Q9SIT3
19	46	47.9	409	10	Q9LV54

20	46	47.9	567	5	Q9NLC9	Q9nlc9 caenorhabdi
21	46	47.9	571	5	Q9NLD0	Q9nld0 caenorhabdi
22	46	47.9	860	3	P87295	P87295 schizosacch
23	46	47.9	8243	5	O96554	O96554 cryptospori
24	45	46.9	256	10	Q08333	Q08333 zea mays
25	44.5	46.4	168	10	P82403	P82403 spinacia ol
26	44.5	46.4	361	1	O57936	O57936 pyrococcus
27	44.5	45.8	24	2	Q9R5S2	Q9r5s2 lactobacill
28	44	45.8	517	10	Q9S792	Q9s792 nicotiana t
29	44	45.8	978	2	Q9PNK9	Q9pnk9 campylobact
30	43	44.8	234	10	Q9LIQ6	Q9liq6 arabidopsi
31	43	44.8	268	2	O69068	O69068 pseudomonas
32	43	44.8	319	2	Q9JYJ3	Q9jyj3 neisseria m
33	43	44.8	342	2	Q9PHT9	Q9pht9 campylobact
34	43	44.8	553	10	Q9AV84	Q9av84 oryza sativ
35	43	44.8	585	5	Q9U4F0	Q9u4f0 drosophila
36	43	44.8	601	10	Q9S7A8	Q9s7a8 arabidopsi
37	43	44.8	695	2	Q9FCR8	Q9fcr8 neisseria m
38	43	44.8	698	2	Q9JPL7	Q9jpl7 neisseria m
39	43	44.8	702	4	Q9P2M4	Q9p2m4 homo sapien
40	43	44.8	715	2	Q9AKU6	Q9aku6 neisseria m
41	43	44.8	990	5	Q9U4F1	Q9u4f1 drosophila
42	43	44.8	1020	5	Q9U1K1	Q9u1k1 drosophila
43	43	44.8	1337	5	Q9Y008	Q9y008 plasmodium
44	43	44.8	1342	5	O97364	O97364 plasmodium
45	43	44.8	1829	5	Q22248	Q22248 caenorhabdi

## ALIGNMENTS

RESULT 1  
Q9N9G9 PRELIMINARY: PRT; 1433 AA.  
AC Q9N9G9;  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE ERYTHROCYTE BINDING PROTEIN (FRAGMENT).  
GN EBP.  
OS Plasmodium reichenowi.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
OX NCBI\_TaxID=5854;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CONGOLESE;  
RA Ozwara H., Kocken C.H.M., Thomas A.W.;  
RT "Molecular characterization of erythrocyte binding protein of  
RT Chimpanzee malaria parasite plasmodium reichenowi.";  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ251848; CAB96159.1; --  
DR InterPro; IPR001596; Pyrophosphatase.  
DR PROSITE; PS00387; PPASE; UNKNOWN\_1.  
FT NON\_TER 1  
FT CHAIN 9 >1433 ERYTHROCYTE BINDING PROTEIN.  
FT NON\_TER 1433 1433  
SQ SEQUENCE 1433 AA; 166648 MW; 733D0CDB6EDC27AE CRC64;

Query Match 100.0%; Score 96; DB 5; Length 1433;  
Best Local Similarity 100.0%; Pred. No. 1.7e-05;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NEREDERTLTKEYEDIVLK 19  
|||||  
Db 1083 NEREDERTLTKEYEDIVLK 1101

RESULT 2  
Q25842 PRELIMINARY: PRT; 1475 AA.  
ID Q25842  
AC Q25842;  
DT 01-NOV-1996 (Tremblrel. 01, Created)

01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DE ERYTHROCYTE BINDING PROTEIN.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN SEQUENCE FROM N.A.  
RP STRAIN=FCC1/HN;  
RC Ma C.L., Yu X.B., Li X.R., Shan Z.X.;  
RA "Conservation and antigenicity of erythrocyte binding antigen 175 in  
RT Plasmodium falciparum isolate FCCI/HN."  
RL Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.  
DR EMBL: AF258781; AAF72186.1;  
SQ SEQUENCE 1435 AA; 167269 MW; FE958828C7794F22 CRC64;

Query Match 92.7%; Score 89; DB 5; Length 1475;  
Best Local Similarity 100.0%; Pred. No. 1.8e-05;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NEREDERTLTKEYEDIVLK 19  
||||| ||||| ||||| |||||

Db 1069 NEREGERTLTKEYEDIVLK 1087

---

RESULT 3  
Q05644 PRELIMINARY; PRT; 1421 AA.

ID Q05644;  
AC Q05644;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE ERYTHROCYTE BINDING ANTIGEN (EBA-175) (FRAGMENT).  
GN EBA-175  
OS Plasmodium falciparum (isolate 3D7), and  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329, 5833;  
RN SEQUENCE FROM N.A.  
RP SPECIES=P. falciparum; STRAIN=ISOLATE 3D7;  
RA Daugherty J.R., Lanar D.E.;  
RL Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.  
[2]  
SEQUENCE OF 795-918 FROM N.A.  
SPECIES=P. falciparum; STRAIN=ISOLATE FCR-3;  
RX MEDLINE=93376008; PubMed=8366884;  
RA Ware L.A., Kain K.C., Sim B.K., Haynes J.D., Baird J.K., Lanar D.E.;  
RT "Two alleles of the 175-kilodalton Plasmodium falciparum erythrocyte  
binding antigen.";  
RL Mol. Biochem. Parasitol. 60:105-109(1993).  
DR EMBL: U32207; AAA75179.1;  
DR EMBL: L07755; AAA02927.1;  
KW Antigen; Polymorphism.  
FT NON\_TER 1421 1421  
SQ SEQUENCE 1421 AA; 165330 MW; 7BB8DDFA07CEE771 CRC64;

Query Match 92.7%; Score 89; DB 5; Length 1421;  
Best Local Similarity 94.7%; Pred. No. 0.00017;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NEREDERTLTKEYEDIVLK 19  
||||| ||||| ||||| |||||

Db 1096 NEREGERTLTKEYEDIVLK 1114

---

RESULT 4  
Q9NG63 PRELIMINARY; PRT; 1435 AA.

ID Q9NG63  
AC Q9NG63;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)

```

RN RP SEQUENCE FROM N.A.
RC STRAIN-2996;
RX MEDLINE=96039602; PubMed=7590185;
RA Rokbi B., Maitre-Wilmotte G., Mazarin V., Fourrichon L., Lissolo L.,
RA Quentin-Millet M.J.;
RT "Variable sequences in a mosaic-like domain of meningococcal tbp2
RT encode immunoreactive epitopes.";
RL FEMS Microbiol. Lett. 132:277-283(1995).
DR EMBL: X88867; CAA61337.1; -.
DR InterPro: IPR001677; Transferrin_bind.
DR Pfam: PF01298; Lipoprotein_5; 1.
FT NON_TER 1
FT NON_TER 217
SQ SEQUENCE 217 AA; 22596 MW; 63A98BAA62D70338 CRC64;

Query Match 50.0%; Score 48; DB 2; Length 217;
Best Local Similarity 50.0%; Pred. No. 21;
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 EREDERTLTKEYEDIVLK 19
DB 183 ERTDEKEIPKEQDIIYVR 200
|| |||: ||: ||||:

RESULT 7
Q9FCS0 PRELIMINARY; PRT; 669 AA.
AC Q9FCS0;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE TRANSFERRIN BINDING PROTEIN B (FRAGMENT).
GN TBPB.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-9094;
RX MEDLINE=20407297; PubMed=10948108;
RA Rokbi B., Renaud-Mongenie G., Mignon M., Danve B., Poncet D.,
RA Chabanel C., Caugant D.A., Quentin-Millet M.J.;
RT "Allelic diversity of the two transferrin binding protein B gene
RT isotypes among a collection of Neisseria meningitidis strains
RT representative of serogroup b disease: Implication for the composition
RT of a recombinant TbpB-based vaccine.";
RL Infect. Immun. 68:4938-4947(2000).
DR EMBL: AJ279557; CAC05591.1; -.
DR InterPro: IPR001677; Transferrin_bind.
DR Pfam: PF01298; Lipoprotein_5; 1.
FT NON_TER 1
FT NON_TER 217
SQ SEQUENCE 669 AA; 72089 MW; 1390E941B9A6E5B4 CRC64;

Query Match 50.0%; Score 48; DB 2; Length 669;
Best Local Similarity 50.0%; Pred. No. 66;
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 EREDERTLTKEYEDIVLK 19
DB 506 ERTDEKEIPKEQDIIYVR 523
|| |||: ||: ||||:

RESULT 8
Q9FCR5 PRELIMINARY; PRT; 669 AA.
AC Q9FCR5;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE TRANSFERRIN BINDING PROTEIN B (FRAGMENT).

```

```

GN TBPB.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SB22;
RX MEDLINE=20407297; PubMed=10948108;
RA Rokbi B., Renaud-Mongenie G., Mignon M., Danve B., Poncet D.,
RA Chabanel C., Caugant D.A., Quentin-Millet M.J.;
RT "Allelic diversity of the two transferrin binding protein B gene
RT isotypes among a collection of Neisseria meningitidis strains
RT representative of serogroup b disease: Implication for the composition
RT of a recombinant TbpB-based vaccine.";
RL Infect. Immun. 68:4938-4947(2000).
DR EMBL: AJ279557; CAC05591.1; -.
DR InterPro: IPR001677; Transferrin_bind.
DR Pfam: PF01298; Lipoprotein_5; 1.
FT NON_TER 1
FT NON_TER 217
SQ SEQUENCE 669 AA; 72161 MW; 01871F185FB741A2 CRC64;

Query Match 50.0%; Score 48; DB 2; Length 669;
Best Local Similarity 50.0%; Pred. No. 66;
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 EREDERTLTKEYEDIVLK 19
DB 506 ERTDEKEIPKEQDIIYVR 523
|| |||: ||: ||||:

RESULT 9
Q51283 PRELIMINARY; PRT; 680 AA.
AC Q51283;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE TRANSFERRIN BINDING PROTEIN B (TRANSFERRIN-BINDING PROTEIN 2)
DE (FRAGMENT).
GN TBPB OR TBP2.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-32794;
RX MEDLINE=97445911; PubMed=9302199;
RA Rokbi B., Mignon M., Caugant D.A., Quentin Millet M.J.;
RT "Heterogeneity of tbpB, the transferrin-binding protein B gene, among
RT serogroup B Neisseria meningitidis strains of the ET-5 complex.";
RL Clin. Diagn. Lab. Immunol. 4:522-529(1997).
RN [2]
RP SEQUENCE OF 16-56 FROM N.A.
RC STRAIN-B15P1.16;
RX MEDLINE=93224009; PubMed=8468000;
RA Wilton J., Ala'Aldeen D., Palmer H.M., Borriello S.P.;
RT "Cloning and partial sequence of transferrin-binding protein 2 of
RT Neisseria meningitidis using a novel method: twin N-terminal PCR.";
RL FEMS Microbiol. Lett. 107:59-66(1993).
DR EMBL: Y09617; CAA70832.1; -.
DR EMBL: X75167; CAA53009.1; -.
DR InterPro: IPR001677; Transferrin_bind.
DR Pfam: PF01298; Lipoprotein_5; 1.
FT NON_TER 1
FT NON_TER 680
SQ SEQUENCE 680 AA; 73442 MW; 1EF098203B3D8282 CRC64;

Query Match 50.0%; Score 48; DB 2; Length 680;
Best Local Similarity 50.0%; Pred. No. 68;
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

```

Mon Feb 4 15:23:43 2002

```

Qy 2 EREDERTLTKEYEDIVLK 19
  || || : || : || :
Db 517 ERTDEKEIPKEQDIVR 534

RESULT 10
Q9AKU7 PRELIMINARY; PRT; 685 AA.
ID Q9AKU7
AC Q9AKU7; 01-JUN-2001 (TREMELREL. 17, Created)
DT 01-JUN-2001 (TREMELREL. 17, Last sequence update)
DT 01-JUN-2001 (TREMELREL. 17, Last annotation update)
DE TRANSFERRIN BINDING PROTEIN B.
GN TBPB.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=92/123;
RA Rokbi B.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; 117806; CAC27780.1;
SQ SEQUENCE 685 AA; 73775 MW; 30ADB39EF34B31A5 CRC64;

Query Match 50.0%; Score 48; DB 2; Length 685;
Best Local Similarity 50.0%; Pred. No. 68;
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 2 EREDERTLTKEYEDIVLK 19
  || || : || : || :
Db 522 ERTDEKEIPKEQDIVR 539

RESULT 11
Q51288 PRELIMINARY; PRT; 689 AA.
ID Q51288
AC Q51288; 01-NOV-1996 (TREMELREL. 01, Created)
DT 01-NOV-1996 (TREMELREL. 01, Last sequence update)
DT 01-JUN-2001 (TREMELREL. 17, Last annotation update)
DE TRANSFERRIN-BINDING PROTEIN 2 PRECURSOR.
GN TBP2.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B283;
RA Legrain M.; Findeli A., Villevall D., Quentin-Millet M., Jacobs E.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; 250732; CAA90599.1;
DR InterPro; IPR001677; Transferrin_bind.
DR Pfam; PF01298; Lipoprotein_5; 1.
KW Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 689 TRANSFERRIN-BINDING PROTEIN 2.
SQ SEQUENCE 689 AA; 74321 MW; 2D1D467569E6A722 CRC64;

Query Match 50.0%; Score 48; DB 2; Length 689;
Best Local Similarity 50.0%; Pred. No. 68;
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 2 EREDERTLTKEYEDIVLK 19
  || || : || : || :
Db 526 ERTDEKEIPKEQDIVR 543

RESULT 12
Q9EXC8 PRELIMINARY; PRT; 697 AA.
ID Q9EXC8

```

```

Q9EXC8;
01-MAR-2001 (TREMELREL. 16, Created)
01-MAR-2001 (TREMELREL. 16, Last sequence update)
01-JUN-2001 (TREMELREL. 17, Last annotation update)
DE TRANSFERRIN BINDING PROTEIN B SUBUNIT (FRAGMENT).
GN TBPB.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z3905;
RA Achtman M., Ende A., Zhu P., Koroleva I.S., Kusecek B., Morelli G.,
  Schuurman I.G.A., Brieske N., Zurth K., Kostyukova N.N.,
  Platonov A.E.;
RT "Clonal groupings associated with successive waves of serogroup A
  meningococcal disease from 1969 to 1997 in Moscow, Russia.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Z3905;
RA Zhu P., Ende A., Brieske N., Morelli G., Popovic T., Riou J.Y.,
  Caugant D.A., Achtman M.;
RT "Microevolution in subgroup III Neisseria meningitidis during three
  pandemics of epidemic meningitis.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Z3905;
RA MEDLINE=20305048; PubMed=10844690;
  Linz B., Schenker M., Achtman M.;
RT "Frequent interspecific genetic exchange between commensal neisseriae
  and Neisseria meningitidis.";
RL Mol. Microbiol. 36:1049-1058(2000).
DR EMBL; AJ276915; CAC21580.1;
DR InterPro; IPR001677; Transferrin_bind.
DR Pfam; PF01298; Lipoprotein_5; 1.
FT NON_TER 1
SQ SEQUENCE 697 AA; 75846 MW; 3D577D2AC8DF129A CRC64;

Query Match 50.0%; Score 48; DB 2; Length 697;
Best Local Similarity 50.0%; Pred. No. 69;
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 2 EREDERTLTKEYEDIVLK 19
  || || : || : || :
Db 538 ERTDEKEIPKEQDIVR 555

RESULT 13
Q9FCR9 PRELIMINARY; PRT; 702 AA.
ID Q9FCR9
AC Q9FCR9; 01-MAR-2001 (TREMELREL. 16, Created)
DT 01-MAR-2001 (TREMELREL. 16, Last sequence update)
DT 01-JUN-2001 (TREMELREL. 17, Last annotation update)
DE TRANSFERRIN BINDING PROTEIN B (FRAGMENT).
GN TBPB.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=92123;
RA MEDLINE=20407297; PubMed=10948108;
  Rokbi B., Renauld-Mongenie G., Mignon M., Danve B., Poncet D.,
  Chabanel C., Caugant D.A., Quentin-Millet M.J.;
RT "Allelic diversity of the two transferrin binding protein B Gene
  isotypes among a collection of Neisseria Meningitidis strains
  representative of serogroup b disease: implication for the composition
  of a recombinant TbpB-based vaccine.";
RL Infect. Immun. 68:4938-4947(2000).

```



DR EMBL; AJ279558; CAC05592.1; -.  
 DR InterPro; IPR001677; Transferrin\_bind.  
 DR Pfam; PF01298; Lipoprotein\_5; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 702 AA; 75796 MW; 14DCB6B64DC13F75 CRC64;

Query Match 50.0%; Score 48; DB 2; Length 702;  
 Best Local Similarity 50.0%; Pred. No. 70;  
 Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 EREDERTLTKEYEDIVLK 19  
 || || : || : || :  
 DB 543 ERTDEKEIPKEQODIVYR 560

## RESULT 14

Q9EXB0 PRELIMINARY; PRT; 707 AA.  
 AC Q9EXB0;  
 DT 01-MAR-2001 (TremBLrel. 16, Created)  
 DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)  
 DE TRANSFERRIN BINDING PROTEIN B SUBUNIT (FRAGMENT).  
 GN TBPB.  
 OS Neisseria meningitidis.  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=487;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-24311;  
 RA Achtman M., Ende A., Zhu P., Koroleva I.S., Kusecek B., Morelli G.,  
 RA Schuurman I.G.A., Brieske N., Zurth K., Kostyukova N.N.,  
 RA Platonov A.E.;  
 RT "Clonal groupings associated with successive waves of serogroup A  
 meningococcal disease from 1969 to 1997 in Moscow, Russia."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-24311;  
 RA Zhu P., Ende A., Brieske N., Morelli G., Popovic T., Riou J.Y.,  
 RA Caugant D.A., Achtman M.;  
 RT "Microevolution in subgroup III Neisseria meningitidis during three  
 pandemics of epidemic meningitis."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-24311;  
 RX MEDLINE-20305048; PubMed-10844690;  
 RA Linz B., Schenker M., Achtman M.;  
 RT "Frequent interspecific genetic exchange between commensal neisseriae  
 and Neisseria meningitidis."  
 RL Mol. Microbiol. 36:1049-1058(2000).  
 DR EMBL; AJ276933; CAC21596.1; -.  
 DR InterPro; IPR001677; Transferrin\_bind.  
 DR Pfam; PF01298; Lipoprotein\_5; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 707 AA; 75729 MW; 9271A060B9DFED04 CRC64;

Query Match 50.0%; Score 48; DB 2; Length 707;  
 Best Local Similarity 50.0%; Pred. No. 70;  
 Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 EREDERTLTKEYEDIVLK 19  
 || || : || : || :  
 DB 542 ERTDEKEIPKEQODIVYR 559

## RESULT 15

Q9AKU8 PRELIMINARY; PRT; 722 AA.  
 ID Q9AKU8  
 AC Q9AKU8;

DT 01-JUN-2001 (TremBLrel. 17, Created)  
 DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)  
 DE TRANSFERRIN BINDING PROTEIN B.  
 GN TBPB.  
 OS Neisseria meningitidis (serogroup B).  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-90/94;  
 RA Rokbi B.;  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Y17805; CAC27779.1; -.  
 SQ SEQUENCE 722 AA; 77921 MW; 0169743E83144B8D CRC64;

Query Match 50.0%; Score 48; DB 2; Length 722;  
 Best Local Similarity 50.0%; Pred. No. 72;  
 Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 EREDERTLTKEYEDIVLK 19  
 || || : || : || :  
 DB 563 ERTDEKEIPKEQODIVYR 580

Search completed: January 29, 2002, 11:12:15  
 Job time: 771 sec

us-09-763-397a-20.rspt

Mon Feb 4 15:23:43 2002

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 10:21:51 : Search time 310.82 Seconds  
(without alignments)  
1.907 Million cell updates/sec

Title: US-09-763-397A-21  
Perfect score: 41  
Sequence: 1 LTPLEELY 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_1101.\*  
1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID58/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID58/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID58/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	41	100.0	8	21	AA70297
2	41	100.0	9	18	AAW35494
3	41	100.0	350	21	AA70278
4	41	100.0	782	11	AAW06991
5	39	95.1	123	21	AAW00310
6	39	95.1	447	12	AAW12388
7	36	87.8	92	21	AAW00240
8	36	87.8	284	22	AAW3892
9	34	82.9	283	20	AAW36933
10	33	80.5	155	22	AAW39296
11	32	78.0	51	21	AAW60904

12	32	78.0	52	21	AAW60903
13	32	78.0	52	21	AAW61365
14	32	78.0	76	22	AAW39585
15	32	78.0	80	21	AAW59375
16	32	78.0	81	21	AAW60029
17	32	78.0	85	22	AAW41371
18	32	78.0	89	21	AAW15329
19	32	78.0	106	21	AAW57874
20	32	78.0	106	21	AAW59661
21	32	78.0	106	21	AAW59662
22	32	78.0	108	21	AAW59374
23	32	78.0	109	21	AAW60028
24	32	78.0	113	21	AAW59373
25	32	78.0	114	21	AAW60027
26	32	78.0	117	21	AAW15328
27	32	78.0	122	21	AAW15327
28	32	78.0	134	21	AAW57873
29	32	78.0	134	21	AAW59660
30	32	78.0	134	21	AAW59961
31	32	78.0	139	21	AAW57872
32	32	78.0	139	21	AAW59659
33	32	78.0	139	21	AAW59960
34	32	78.0	414	20	AAW34996
35	32	78.0	416	17	AAW06795
36	32	78.0	539	21	AAW36614
37	32	78.0	556	21	AAW36613
38	32	78.0	568	21	AAW36612
39	31	75.6	75	18	AAW20395
40	31	75.6	159	18	AAW20665
41	31	75.6	271	21	AAW19619
42	31	75.6	371	17	AAW83015
43	31	75.6	371	22	AAW05583
44	31	75.6	2273	19	AAW70398
45	30	73.2	18	22	AAW98058

#### ALIGNMENTS

RESULT 1

AA70297  
ID AA70297 standard; peptide; 8 AA.

AC AA70297;

DT 06-JUN-2000 (first entry)

XX Plasmodium falciparum RAP-1 antigenic epitope, p545.

XX Recombinant protein; CDC/NIAID/VAC-1; multivalent; malaria; vaccine;  
KW T-cell epitope; tetanus toxoid; antigenic epitope; treatment;  
KW circumsporozoite protein; CSP; sporozoite surface protein-2; SSP-2;  
KW liver stage antigen-1; LSA-1; merozoite surface protein-1; MSP-1; MSP-2;  
KW apical membrane antigen-1; AMA-1; erythrocyte binding antigen-175;  
KW EBA-175; rhoptry associated protein-1; RAP-1; gamete specific antigen;  
KW Pf27; antiparasitic; prevention; anti-CDC/NIAID/VAC-1 antibody.

OS Plasmodium falciparum.

XX WO200011179-A1.

XX PD 02-MAR-2000.

XX PF 19-AUG-1999; 99WO-US18869.

XX PR 21-AUG-1998; 98US-0097703.

XX PA (NAIM-) NAT INST IMMUNOLOGY.  
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX PI Lal AA, Shi YP, Hasnain SE;

XX WPI; 2000-237654/20.

XX Novel recombinant protein as vaccine for treating malarial infection  
PT comprises antigenic peptides obtained from different stages of  
XX plasmodium falciparum life cycle  
XX  
XX Claim 2; Page 17; 52pp; English.  
XX  
XX The present sequence is the antigenic epitope P545, derived from  
CC rhoptry associated protein-1 (RAP-1) of the asexual blood stage of  
CC Plasmodium falciparum. It is used in the construction of recombinant  
CC protein CDC/NIIMALVAC-1, which is a multivalent, multistage malarial  
CC vaccine. The recombinant protein comprises, melittin signal peptide,  
CC (His)6 tag, T-cell epitope from tetanus toxoid and 21 antigenic epitopes  
CC from circumsporozoite protein (CSP), sporozoite surface protein-2  
CC (SSP-2), liver stage antigen-1 (LSA-1), merozoite surface protein-1  
CC (MSP-1), MSP-2, apical membrane antigen-1 (AMA-1), erythrocyte binding  
CC antigen-175 (EBA-175), rhoptry associated protein-1 (RAP-1) and gamete  
CC specific antigen, Pf27. These epitopes were obtained at different stages  
CC of the life cycle of P. falciparum. CDC/NIIMALVAC-1 vaccine has  
CC antiparasitic activity and can be used for treatment and prevention of  
CC malarial infections. Anti-CDC/NIIMALVAC-1 antibodies can be used for  
CC detecting P. falciparum in biological samples.

SQ Sequence 8 AA;  
Query Match 100.0%; Score 41; DB 21; Length 8;  
Best Local Similarity 100.0%; Pred. No. 4.3e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTPLEELY 8  
DB 1 ltpleely 8

RESULT 2  
AAW35494  
ID AAW35494 standard; peptide; 9 AA.

XX AAW35494;  
XX  
XX 22-APR-1998 (first entry)  
XX  
XX Band-3 peptide SEQ ID NO:22 from WO9738011.  
XX  
XX T-cell stimulatory peptide; immunogen; non-dendritic; carrier; tumour;  
XX scaffold; inhibition; metastasis; wound healing; solid phase.  
XX  
XX Unidentified.

XX  
XX WO9738011-A1.  
XX  
XX 16-OCT-1997.  
XX  
XX 03-APR-1997; 97WO-DE00146.  
XX  
XX 03-APR-1996; 96DK-0000398.  
XX  
XX (PEPR-) PEPRESEARCH AS.  
XX  
XX Heegaard PMH, Jakobsen PH;  
XX  
XX WPI; 1997-512645/47.  
XX

XX Non-dendritic peptide carrier linked to a solid phase - useful as a  
PT diagnostic agent and as a scaffold for production of chemical  
PT derivatives  
XX

XX Example 5; Page 92; 262pp; English.

XX A non-dendritic peptide carrier (A) has been developed which is coupled  
CC through a linker to a solid phase, forming a complex of (A)-solid phase.  
XX Where (A) comprises 10-50 amino acids capable of forming a secondary

CC structure in a benign buffer after liberation from the solid phase, and  
CC further the (A)-solid phase complex comprises an immunogenic substance  
CC and/or an immune mediator coupled on (A). The present invention. An  
CC represents a peptide used in an example for the present invention. An  
CC (A)-solid phase complex can be used as a scaffold for the production of  
CC chemical derivatives, characterised by covalently attaching molecules at  
CC attachment points. Alternatively (A) is used as a scaffold-peptide for  
CC the incorporation into an immunostimulating complex (iscom) resulting an  
CC (A)-iscom complex which is used for the chemical coupling of antigenic  
CC substances in an aqueous solution by conjugation. (A) derivatised with  
CC one or more peptides having fibronectin-, laminin- or vitronectin-like  
CC binding activities can be used for the promotion of cell-attachment to  
CC plastic surfaces, in particular to inhibit tumour growth and metastasis,  
CC and for promotion of wound healing. Also a derivatised (A) can be used  
CC for the selection of specifically-binding aptamers or as a diagnostic  
CC agent. Such diagnostic (A) molecules could be used to detect molecules  
CC derived from or indicative of pregnancy or of a disease, such as an  
CC infectious, autoimmune or cancerous disease.

XX Sequence 9 AA;

Query Match 100.0%; Score 41; DB 18; Length 9;  
Best Local Similarity 100.0%; Pred. No. 4.3e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTPLEELY 8  
DB 1 ltpleely 8

RESULT 3  
AAW70278  
ID AAY70278 standard; Protein; 350 AA.

XX AAY70278;  
XX  
XX 06-JUN-2000 (first entry)  
XX  
XX Recombinant vaccine CDC/NIIMALVAC-1.  
XX  
XX Recombinant protein; CDC/NIIMALVAC-1; multivalent; malaria; vaccine;  
XX T-cell epitope; tetanus toxoid; antigenic epitope; treatment;  
XX circumsporozoite protein; CSP; sporozoite surface protein-2; SSP-2;  
XX liver stage antigen-1; LSA-1; merozoite surface protein-1; MSP-1; MSP-2;  
XX apical membrane antigen-1; AMA-1; erythrocyte binding antigen-175;  
XX EBA-175; rhoptry associated protein-1; RAP-1; gamete specific antigen;  
XX Pf27; antiparasitic; prevention; anti-CDC/NIIMALVAC-1 antibody;  
XX honey bee.

XX Chimeric - Apis sp.  
XX Chimeric - Clostridium tetani.  
XX Chimeric - Plasmodium falciparum.

XX Key Location/Qualifiers  
XX Peptide 1..22  
XX /label= Melittin\_signal\_peptide  
XX /note= "Derived from Honey bee"  
XX Protein 23..350  
XX /label= Mature\_CDC/NIIMALVAC-1  
XX /note= "Recombinant multivalent malarial vaccine"

XX WO200011179-A1.

XX 02-MAR-2000.  
XX 19-AUG-1999; 99WO-US18869.  
XX 21-AUG-1998; 98US-0097703.  
XX (NALM-) NAT INST IMMUNOLOGY.  
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Lal AA, Shi YP, Hasnain SE;  
 XX WPI: 2000-237654/20.  
 DR N-PSDB; AAZ51336.  
 XX  
 PT Novel recombinant protein as vaccine for treating malarial infection  
 PT comprises antigenic peptides obtained from different stages of  
 PT plasmodium falciparum life cycle -  
 XX  
 PS Claim 3; Page 43-44; 52pp; English.  
 XX  
 CC The present sequence is that of recombinant protein CDC/NIIMALVAC-1,  
 CC which is a multivalent, multistage malarial vaccine. The recombinant  
 CC protein comprises, melittin signal peptide, (His)6 tag, T-cell epitope  
 CC from tetanus toxoid and 21 antigenic epitopes from circumsporozoite  
 CC protein (CSP), sporozoite surface protein-2 (SSP-2), liver stage  
 CC antigen-1 (LSA-1), merozoite surface protein-1 (MSP-1), MSP-2, apical  
 CC membrane antigen-1 (AMA-1), erythrocyte binding antigen-175 (EBA-175),  
 CC rhoptry associated protein-1 (RAP-1) and gamete specific antigen, Pf27.  
 CC These epitopes were obtained at different stages of the life cycle of  
 CC Plasmodium falciparum. CDC/NIIMALVAC-1 vaccine has antiparasitic  
 CC activity and can be used for treatment and prevention of malarial  
 CC infections. Anti-CDC/NIIMALVAC-1 antibodies can be used for detecting  
 CC P. falciparum in biological samples.  
 XX  
 SQ Sequence 350 AA;  
 Query Match 100.0%; Score 41; DB 21; Length 350;  
 Best Local Similarity 100.0%; Pred. No. 3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LTPLEELY 8  
 Db 176 ltpleely 183  
 |||||  
 RESULT 4  
 AAR06991  
 ID AAR06991 standard; protein; 782 AA.  
 XX  
 AC AAR06991;  
 XX  
 DT 16-JAN-1991 (first entry)  
 XX  
 DE Polypeptide antigenic to rhostry organelles of Plasmodium.  
 XX  
 KW Malaria; vaccine; rhostry organelles.  
 XX  
 OS Plasmodium falciparum.  
 XX  
 PN EP388738-A.  
 XX  
 PD 26-SEP-1990.  
 XX  
 PF 09-MAR-1990; 90EP-0104561.  
 XX  
 PR 22-AUG-1989; 89GB-0019064.  
 PR 14-MAR-1989; 89GB-0005857.  
 XX  
 PA (HOFF ) HOFFMANN-LA ROCHE AG.  
 XX  
 PI Ridley RG, Scaife JG;  
 XX  
 DR WPI: 1990-291721/39.  
 DR N-PSDB; AAQ06000.  
 XX  
 PT Antigenic polypeptide and DNA encoding it - having a determinant  
 PT cross reactive with those on the rhostry organelles of the  
 PT merozoite form of the malaria parasite P falciparum.  
 XX  
 PS Claim 1; Fig 2a-b; 29pp; English.  
 XX

CC Gene product may be isolated from a transformed E.coli (Y1088)  
 CC expression system using plasmid pMC9, and may then be used for  
 CC immunisation against malaria. The product may also be used  
 CC diagnostically to detect Abs directed against the parasite.  
 XX  
 SQ Sequence 782 AA;  
 Query Match 100.0%; Score 41; DB 11; Length 782;  
 Best Local Similarity 100.0%; Pred. No. 7.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LTPLEELY 8  
 Db 202 ltpleely 209  
 |||||  
 RESULT 5  
 AAG00310  
 ID AAG00310 standard; Protein; 123 AA.  
 XX  
 AC AAG00310;  
 XX  
 DT 06-OCT-2000 (first entry)  
 XX  
 DE Human secreted protein, SEQ ID NO: 4391.  
 XX  
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1033401-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 21-FEB-2000; 2000EP-0200610.  
 XX  
 PR 26-FEB-1999; 99US-0122487.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 PI Dumas Milne Edwards J, Duclert A, Giordano J;  
 XX  
 DR WPI: 2000-500381/45.  
 DR N-PSDB; AAC00316.  
 XX  
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX  
 PS Claim 13; SEQ ID 4391; 71pp + CD-ROM; English.  
 XX  
 CC The present sequence is a polypeptide encoded by one of a large number  
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30  
 CC different tissues. EST sequences usually correspond mainly to the 3'  
 CC untranslated region (UTR) of the mRNA because they are often obtained  
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
 CC those cases where longer cDNA sequences have been obtained, the full 5'  
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
 CC ends and can therefore be used to obtain full length cDNAs and genomic  
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
 CC chromosome mapping procedures. They are used to obtain upstream  
 CC regulatory sequences and to design expression and secretion vectors.  
 XX  
 SQ Sequence 123 AA;  
 Query Match 95.1%; Score 39; DB 21; Length 123;  
 Best Local Similarity 87.5%; Pred. No. 2.5;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

us-09-763-397a-21.rag

Mon Feb 4 15:23:43 2002

```

QY 1 LTPLEELY 8
Db 46 itpleely 53

RESULT 6
AAR12388
ID AAR12388 standard; Protein; 447 AA.
XX
XX
AC AAR12388;
XX
XX 29-AUG-1991 (first entry)
DT
DE smg p25A GDP Dissociation Inhibitory Protein.
XX
XX G-protein; Guanosine diphosphate; GDP; GDI.
XX
XX JP03113000-A.
PN
XX 14-MAY-1991.
PD
XX
XX 26-SEP-1989; 89JP-0249441.
PE
XX 26-SEP-1989; 89JP-0249441.
PR
XX (MITU ) MITSUBISHI KASEI CORP.
PA
XX WPI; 1991-183251/25.
XX
XX N-PSDB; AAQ11984.
DR
XX
XX New protein for controlling levels of GTP-binding protein -
PT using binding inhibition, by binding to G-protein to inhibit GDP
PT GDP dissociation to GTP
XX
XX Claim 3; Page 1; 12pp; Japanese.
PS
XX The protein is a GDP Dissociation Inhibitory Protein (GDI) of mol.
XX wt. 50-60kD which binds to low mol. wt. G-protein smg p25A.
CC
XX
XX Sequence 447 AA;
SQ

Query Match 95.1%; Score 39; DB 12; Length 447;
Best Local Similarity 87.5%; Pred. No. 9.8;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTPLEELY 8
Db 46 itpleely 53

RESULT 7
AAG00240
ID AAG00240 standard; Protein; 92 AA.
XX
XX
AC AAG00240;
XX
XX 06-OCT-2000 (first entry)
DT
DE Human secreted protein, SEQ ID NO: 4321.
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping.
XX
XX Homo sapiens.
OS
XX EP1033401-A2.
XX
XX 06-SEP-2000.
PD
XX
XX 21-FEB-2000; 2000EP-0200610.
PF
XX
XX
XX

```

```

26-FEB-1999; 99US-0122487.
(GEST ) GENSET.
Dumas Milne Edwards J, Duclert A, Giordano J;
WPI; 2000-500381/45.
N-PSDB; AAC00246.
New nucleic acid that is a 5' expressed sequence tag (5' EST) for
obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
diagnostic, forensic, gene therapy and chromosome mapping procedures -
Claim 13; SEQ ID 4321; 71pp + CD-ROM; English.
The present sequence is a polypeptide encoded by one of a large number
of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
were prepared from total human RNAs or polyA+ RNAs derived from 30
different tissues. EST sequences usually correspond mainly to the 3'
untranslated region (UTR) of the mRNA because they are often obtained
from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
isolating cDNA sequences derived from the 5' ends of mRNAs and even in
those cases where longer cDNA sequences have been obtained, the full 5'
UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
ends and can therefore be used to obtain full length cDNAs and genomic
DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
chromosome mapping procedures. They are used to obtain upstream
regulatory sequences and to design expression and secretion vectors.
XX Sequence 92 AA;
SQ

Query Match 87.8%; Score 36; DB 21; Length 92;
Best Local Similarity 75.0%; Pred. No. 7.2;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTPLEELY 8
Db 46 itpleely 53

RESULT 8
AAG73892
ID AAG73892 standard; Protein; 284 AA.
XX
XX
AC AAG73892;
XX
XX 03-SEP-2001 (first entry)
DT
DE Human colon cancer antigen protein SEQ ID NO:4656.
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma; chromosome 2.
XX
XX Homo sapiens.
OS
XX WO200122920-A2.
XX
XX 05-APR-2001.
PD
XX 28-SEP-2000; 2000WO-US26524.
XX
XX 29-SEP-1999; 99US-0157137.
PR
XX 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX WPI; 2001-235357/24.
XX
XX N-PSDB; AAH33323.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT

```

PT useful for preventing, diagnosing and/or treating colorectal cancers -  
 PS Claim 11; Page 6456-6458; 9803pp; English.  
 XX  
 CC AAH32943 to AAH37195 and AAH73514 to AAH77788 represent human colon  
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
 CC the proteins are collectively known as colon cancer antigens. The colon  
 CC cancer antigens have cytostatic activity and can be used in gene  
 CC therapy and vaccine production. N and P may be used in the prevention,  
 CC diagnosis and treatment of diseases associated with inappropriate P  
 CC expression. For example, N and P may be used to treat disorders  
 CC associated with decreased expression by rectifying mutations or deletions  
 CC in a patient's genome that affect the activity of P by expressing  
 CC inactive proteins or to supplement the patients own production of P.  
 CC Additionally, N may be used to produce the colon cancer-associated Ps,  
 CC by inserting the nucleic acids into a host cell and culturing the cell  
 CC to express the proteins. N and P can be used in the prevention, diagnosis  
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
 CC and AAH77789 represent sequences used in the exemplification of the  
 CC present invention.  
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
 CC missing at time of publication, meaning no sequences are present for  
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.  
 XX  
 SQ Sequence 284 AA;

Query Match 87.8%; Score 36; DB 22; Length 284;  
 Best Local Similarity 75.0%; Pred. No. 24;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTPLEELY 8  
 :|||:|  
 Db 104 ltpledly 111

RESULT 9  
 AAH36933  
 ID AAH36933 standard; Protein; 283 AA.

XX AAH36933;

XX 07-OCT-1999 (first entry)

XX Chlamydia trachomatis secreted protein.

XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
 KW paratrachoma; inclusion conjunctivitis; genital disease; perihhepatitis;  
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
 KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.

OS Chlamydia trachomatis.

XX WO9928475-A2.

XX 10-JUN-1999.

XX 27-NOV-1998; 98WO-IB01939.

XX 04-NOV-1998; 98US-0107077.

PR 28-NOV-1997; 97FR-0015041.

XX 17-DEC-1997; 97FR-0016034.

PA (GEST ) GENSET.

XX Griffiths R;

XX WPI; 1999-371125/31.

XX Genome sequence of Chlamydia trachomatis

PS Disclosure; Page 788-789; 1755pp; English.

XX

CC AAH36754-Y37949 are encoded by open reading frames (ORFs) of the genome  
 CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as  
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences  
 CC can also be used to control growth of the microorganism. Chlamydia  
 CC trachomatis is responsible for a large number of diseases, e.g. eye  
 CC diseases such as conventional trachoma, nonendemic trachoma,  
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as  
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,  
 CC perihhepatitis, bartholinitis; pneumopathy in breast feeding infants;  
 CC and venereal lymphogranulomatosis. The polypeptides of the invention  
 CC may be of use in treating these diseases.

XX SQ Sequence 283 AA;

Query Match 82.9%; Score 34; DB 20; Length 283;  
 Best Local Similarity 75.0%; Pred. No. 59;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LTPLEELY 8  
 :|||:|  
 Db 82 ltpienly 89

RESULT 10  
 AAH39296  
 ID AAH39296 standard; Protein; 155 AA.

XX AAH39296;

XX 22-OCT-2001 (first entry)

XX Human polypeptide SEQ ID NO 2441.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Dmanac RT;

XX WPI; 2001-442253/47.

DR N-PSDB; AA158452.

XX Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -

XX Example 4; SEQ ID NO 2441; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and

Mon Feb 4 15:23:43 2002

CC the encoded polypeptides (AAM38642-RAA42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
CC  
SQ Sequence 155 AA;

Query Match 80.5%; Score 33; DB 22; Length 155;  
Best Local Similarity 75.0%; Pred. No. 49;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LTPLEELY 8  
      ||||:|  
Db 137 ltplqey 144

RESULT 11  
AAG60904  
ID AAG60904 standard; Protein; 51 AA.  
XX  
AC AAG60904;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 78941.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 04-MAY-1999; 99US-0132048.  
PR 05-MAY-1999; 99US-0132407.  
PR 06-MAY-1999; 99US-0132484.  
PR 07-MAY-1999; 99US-0132486.  
PR 11-MAY-1999; 99US-0132863.  
PR 99US-0134256.

PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 21-MAY-1999; 99US-0135629.  
PR 24-MAY-1999; 99US-0136021.  
PR 25-MAY-1999; 99US-0136392.  
PR 27-MAY-1999; 99US-0136782.  
PR 28-MAY-1999; 99US-0137222.  
PR 01-JUN-1999; 99US-0137528.  
PR 03-JUN-1999; 99US-0137502.  
PR 04-JUN-1999; 99US-0137724.  
PR 07-JUN-1999; 99US-0138094.  
PR 08-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 10-JUN-1999; 99US-0139119.  
PR 14-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139780.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 20-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 21-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 22-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145216.



PR 23-JUL-1999;	99US-0145224.
PR 26-JUL-1999;	99US-0145276.
PR 27-JUL-1999;	99US-0145913.
PR 27-JUL-1999;	99US-0145918.
PR 27-JUL-1999;	99US-0145919.
PR 28-JUL-1999;	99US-0145951.
PR 02-AUG-1999;	99US-0146386.
PR 02-AUG-1999;	99US-0146388.
PR 02-AUG-1999;	99US-0146389.
PR 03-AUG-1999;	99US-0147038.
PR 04-AUG-1999;	99US-0147204.
PR 04-AUG-1999;	99US-0147302.
PR 05-AUG-1999;	99US-0147192.
PR 05-AUG-1999;	99US-0147260.
PR 06-AUG-1999;	99US-0147303.
PR 06-AUG-1999;	99US-0147416.
PR 09-AUG-1999;	99US-0147493.
PR 09-AUG-1999;	99US-0147935.
PR 10-AUG-1999;	99US-0148171.
PR 11-AUG-1999;	99US-0148319.
PR 12-AUG-1999;	99US-0148341.
PR 13-AUG-1999;	99US-0148565.
PR 13-AUG-1999;	99US-0148684.
PR 16-AUG-1999;	99US-0149368.
PR 17-AUG-1999;	99US-0149175.
PR 18-AUG-1999;	99US-0149426.
PR 20-AUG-1999;	99US-0149722.
PR 20-AUG-1999;	99US-0149723.
PR 20-AUG-1999;	99US-0149929.
PR 23-AUG-1999;	99US-0149902.
PR 23-AUG-1999;	99US-0149930.
PR 25-AUG-1999;	99US-0150556.
PR 26-AUG-1999;	99US-0150884.
PR 27-AUG-1999;	99US-0151065.
PR 27-AUG-1999;	99US-0151066.
PR 27-AUG-1999;	99US-0151080.
PR 30-AUG-1999;	99US-0151303.
PR 31-AUG-1999;	99US-0151438.
PR 01-SEP-1999;	99US-0151930.
PR 07-SEP-1999;	99US-0152363.
PR 10-SEP-1999;	99US-0153070.
PR 13-SEP-1999;	99US-0153758.
PR 15-SEP-1999;	99US-0154018.
PR 16-SEP-1999;	99US-0154039.
PR 20-SEP-1999;	99US-0154779.
PR 22-SEP-1999;	99US-0155119.
PR 23-SEP-1999;	99US-0155486.
PR 24-SEP-1999;	99US-0155659.
PR 28-SEP-1999;	99US-0156458.
PR 29-SEP-1999;	99US-0156596.
PR 04-OCT-1999;	99US-0157117.
PR 05-OCT-1999;	99US-0157753.
PR 06-OCT-1999;	99US-0157865.
PR 07-OCT-1999;	99US-0158029.
PR 08-OCT-1999;	99US-0158232.
PR 12-OCT-1999;	99US-0158369.
PR 13-OCT-1999;	99US-0159293.
PR 13-OCT-1999;	99US-0159294.
PR 13-OCT-1999;	99US-0159295.
PR 14-OCT-1999;	99US-0159329.
PR 14-OCT-1999;	99US-0159330.
PR 14-OCT-1999;	99US-0159331.
PR 14-OCT-1999;	99US-0159637.
PR 14-OCT-1999;	99US-0159638.
PR 18-OCT-1999;	99US-0159584.
PR 21-OCT-1999;	99US-0160741.
PR 21-OCT-1999;	99US-0160767.
PR 21-OCT-1999;	99US-0160768.
PR 21-OCT-1999;	99US-0160770.
PR 21-OCT-1999;	99US-0160814.
PR 21-OCT-1999;	99US-0160815.
PR 22-OCT-1999;	99US-0160980.
PR 22-OCT-1999;	99US-0160981.
PR 23-JUL-1999;	99US-0160989.
PR 25-OCT-1999;	99US-0161404.
PR 25-OCT-1999;	99US-0161405.
PR 25-OCT-1999;	99US-0161406.
PR 26-OCT-1999;	99US-0161359.
PR 26-OCT-1999;	99US-0161360.
PR 26-OCT-1999;	99US-0161361.
PR 28-OCT-1999;	99US-0161920.
PR 28-OCT-1999;	99US-0161992.
PR 28-OCT-1999;	99US-0161993.
PR 29-OCT-1999;	99US-0162142.
Query Match 78.0%; Score 32; DB 21; Length 51;	
Best Local Similarity 100.0%; Pred. No. 24;	
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy 3 PLEELY 8	
Db 34 pleely 39	
RESULT 12	
AAG60903	
ID AAG60903 standard; Protein: 52 AA.	
XX AAG60903;	
XX 18-OCT-2000 (first entry)	
XX Arabidopsis thaliana protein fragment SEQ ID NO: 78939.	
XX Protein identification; signal transduction pathway; metabolic pathway;	
XX hybridisation assay; genetic mapping; gene expression control; promoter;	
XX termination sequence.	
XX Arabidopsis thaliana.	
XX EP1033405-A2.	
PD 06-SEP-2000.	
XX 25-FEB-2000; 2000EP-0301439.	
XX 25-FEB-1999;	99US-0121825.
PR 05-MAR-1999;	99US-0123180.
PR 09-MAR-1999;	99US-0123548.
PR 23-MAR-1999;	99US-0123788.
PR 25-MAR-1999;	99US-0126264.
PR 29-MAR-1999;	99US-0126785.
PR 01-APR-1999;	99US-0127462.
PR 08-APR-1999;	99US-0128234.
PR 16-APR-1999;	99US-0128714.
PR 19-APR-1999;	99US-0129845.
PR 21-APR-1999;	99US-0130077.
PR 23-APR-1999;	99US-0130449.
PR 23-APR-1999;	99US-0130510.
PR 28-APR-1999;	99US-0130891.
PR 30-APR-1999;	99US-0131449.
PR 30-APR-1999;	99US-0132048.
PR 04-MAY-1999;	99US-0132407.
PR 05-MAY-1999;	99US-0132484.
PR 06-MAY-1999;	99US-0132485.
PR 06-MAY-1999;	99US-0132486.
PR 07-MAY-1999;	99US-0132487.
PR 11-MAY-1999;	99US-0132863.
PR 14-MAY-1999;	99US-0134256.
PR 14-MAY-1999;	99US-0134218.
PR 14-MAY-1999;	99US-0134219.
PR 14-MAY-1999;	99US-0134221.
PR 14-MAY-1999;	99US-0134370.
PR 18-MAY-1999;	99US-0134768.
PR 19-MAY-1999;	99US-0134941.
PR 20-MAY-1999;	99US-0135124.

us-09-763-397a-21.rag

Mon Feb 4 15:23:43 2002

PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0138021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137724.  
PR 07-JUN-1999; 99US-0138094.  
PR 08-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 10-JUN-1999; 99US-0139119.  
PR 14-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 16-JUN-1999; 99US-0139454.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 18-JUN-1999; 99US-0139817.  
PR 21-JUN-1999; 99US-0139899.  
PR 22-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 23-JUN-1999; 99US-0140699.  
PR 24-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140991.  
PR 29-JUN-1999; 99US-0141287.  
PR 30-JUN-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 01-JUL-1999; 99US-0142055.  
PR 02-JUL-1999; 99US-0142390.  
PR 06-JUL-1999; 99US-0142803.  
PR 08-JUL-1999; 99US-0142920.  
PR 09-JUL-1999; 99US-0142977.  
PR 12-JUL-1999; 99US-0143542.  
PR 13-JUL-1999; 99US-0143624.  
PR 14-JUL-1999; 99US-0144005.  
PR 15-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 16-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 19-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 20-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 21-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 23-JUL-1999; 99US-0145276.  
PR 26-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 06-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 09-AUG-1999; 99US-0148171.  
PR 10-AUG-1999; 99US-0148319.  
PR 11-AUG-1999; 99US-0148341.  
PR 12-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 13-AUG-1999; 99US-0149368.  
PR 16-AUG-1999; 99US-0149175.  
PR 17-AUG-1999; 99US-0149426.  
PR 18-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 26-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 01-SEP-1999; 99US-0152363.  
PR 07-SEP-1999; 99US-0153070.  
PR 10-SEP-1999; 99US-0153758.  
PR 13-SEP-1999; 99US-0154018.  
PR 15-SEP-1999; 99US-0154039.  
PR 16-SEP-1999; 99US-0154779.  
PR 20-SEP-1999; 99US-0155139.  
PR 22-SEP-1999; 99US-0155486.  
PR 23-SEP-1999; 99US-0155659.  
PR 24-SEP-1999; 99US-0156458.  
PR 28-SEP-1999; 99US-0156596.  
PR 29-SEP-1999; 99US-0157117.  
PR 04-OCT-1999; 99US-0157753.  
PR 05-OCT-1999; 99US-0157865.  
PR 06-OCT-1999; 99US-0158029.  
PR 07-OCT-1999; 99US-0158232.  
PR 08-OCT-1999; 99US-0158369.  
PR 12-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 21-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.

PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 78.0%; Score 32; DB 21; Length 52;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 PLEELY 8  
      |||||  
Db 34 pleely 39

RESULT 13  
AAG61365  
ID AAG61365 standard; Protein; 52 AA.  
XX AC AAG61365;  
XX DT 18-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 79587.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134250.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.



Query Match 78.0%; Score 32; DB 22; Length 76;  
Best Local Similarity 75.0%; Pred. No. 36;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LTPLEELY 8  
| ||:||||  
Db 46 lrpldely 53

## RESULT 15

AAG59375  
ID AAG59375 standard; Protein: 80 AA.

XX AC AAG59375;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 76796.

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

XX PR 05-MAR-1999; 99US-0123180.

XX PR 09-MAR-1999; 99US-0123548.

XX PR 23-MAR-1999; 99US-0125788.

XX PR 25-MAR-1999; 99US-0126264.

XX PR 29-MAR-1999; 99US-0126785.

XX PR 01-APR-1999; 99US-0127462.

XX PR 08-APR-1999; 99US-0128234.

XX PR 16-APR-1999; 99US-0128714.

XX PR 19-APR-1999; 99US-0129845.

XX PR 21-APR-1999; 99US-0130077.

XX PR 23-APR-1999; 99US-0130449.

XX PR 28-APR-1999; 99US-0130891.

XX PR 30-APR-1999; 99US-0131449.

XX PR 04-MAY-1999; 99US-0132048.

XX PR 05-MAY-1999; 99US-0132407.

XX PR 06-MAY-1999; 99US-0132484.

XX PR 07-MAY-1999; 99US-0132486.

XX PR 11-MAY-1999; 99US-0132487.

XX PR 14-MAY-1999; 99US-0132566.

XX PR 14-MAY-1999; 99US-0134218.

XX PR 14-MAY-1999; 99US-0134219.

XX PR 14-MAY-1999; 99US-0134221.  
XX PR 14-MAY-1999; 99US-0134370.  
XX PR 18-MAY-1999; 99US-0134768.  
XX PR 19-MAY-1999; 99US-0134941.  
XX PR 20-MAY-1999; 99US-0135124.  
XX PR 21-MAY-1999; 99US-0135153.  
XX PR 24-MAY-1999; 99US-0135629.  
XX PR 25-MAY-1999; 99US-0136021.  
XX PR 27-MAY-1999; 99US-0136392.  
XX PR 28-MAY-1999; 99US-0136782.  
XX PR 01-JUN-1999; 99US-0137222.  
XX PR 03-JUN-1999; 99US-0137528.  
XX PR 04-JUN-1999; 99US-0137502.  
XX PR 07-JUN-1999; 99US-0137724.  
XX PR 08-JUN-1999; 99US-0138094.  
XX PR 10-JUN-1999; 99US-0138540.  
XX PR 10-JUN-1999; 99US-0138847.  
XX PR 14-JUN-1999; 99US-0139119.  
XX PR 16-JUN-1999; 99US-0139452.  
XX PR 16-JUN-1999; 99US-0139453.  
XX PR 17-JUN-1999; 99US-0139492.  
XX PR 18-JUN-1999; 99US-0139454.  
XX PR 18-JUN-1999; 99US-0139455.  
XX PR 18-JUN-1999; 99US-0139456.  
XX PR 18-JUN-1999; 99US-0139457.  
XX PR 18-JUN-1999; 99US-0139458.  
XX PR 18-JUN-1999; 99US-0139459.  
XX PR 18-JUN-1999; 99US-0139460.  
XX PR 18-JUN-1999; 99US-0139461.  
XX PR 18-JUN-1999; 99US-0139462.  
XX PR 18-JUN-1999; 99US-0139463.  
XX PR 18-JUN-1999; 99US-0139463.  
XX PR 18-JUN-1999; 99US-0139750.  
XX PR 18-JUN-1999; 99US-0139763.  
XX PR 21-JUN-1999; 99US-0139817.  
XX PR 22-JUN-1999; 99US-0139899.  
XX PR 23-JUN-1999; 99US-0140353.  
XX PR 23-JUN-1999; 99US-0140354.  
XX PR 24-JUN-1999; 99US-0140695.  
XX PR 28-JUN-1999; 99US-0140823.  
XX PR 29-JUN-1999; 99US-0140991.  
XX PR 30-JUN-1999; 99US-0141287.  
XX PR 01-JUL-1999; 99US-0141842.  
XX PR 01-JUL-1999; 99US-0142154.  
XX PR 02-JUL-1999; 99US-0142055.  
XX PR 06-JUL-1999; 99US-0142390.  
XX PR 08-JUL-1999; 99US-0142803.  
XX PR 09-JUL-1999; 99US-0142920.  
XX PR 12-JUL-1999; 99US-0142977.  
XX PR 13-JUL-1999; 99US-0143542.  
XX PR 14-JUL-1999; 99US-0143624.  
XX PR 15-JUL-1999; 99US-0144005.  
XX PR 16-JUL-1999; 99US-0144085.  
XX PR 16-JUL-1999; 99US-0144086.  
XX PR 19-JUL-1999; 99US-0144325.  
XX PR 19-JUL-1999; 99US-0144331.  
XX PR 19-JUL-1999; 99US-0144332.  
XX PR 19-JUL-1999; 99US-0144333.  
XX PR 19-JUL-1999; 99US-0144334.  
XX PR 19-JUL-1999; 99US-0144335.  
XX PR 20-JUL-1999; 99US-0144352.  
XX PR 20-JUL-1999; 99US-0144632.  
XX PR 20-JUL-1999; 99US-0144884.  
XX PR 21-JUL-1999; 99US-0144814.  
XX PR 21-JUL-1999; 99US-0145086.  
XX PR 21-JUL-1999; 99US-0145088.  
XX PR 22-JUL-1999; 99US-0145085.  
XX PR 22-JUL-1999; 99US-0145087.  
XX PR 22-JUL-1999; 99US-0145089.  
XX PR 22-JUL-1999; 99US-0145192.  
XX PR 23-JUL-1999; 99US-0145145.  
XX PR 23-JUL-1999; 99US-0145218.  
XX PR 23-JUL-1999; 99US-0145218.  
XX PR 23-JUL-1999; 99US-0145224.  
XX PR 26-JUL-1999; 99US-0145276.  
XX PR 27-JUL-1999; 99US-0145913.  
XX PR 27-JUL-1999; 99US-0145918.  
XX PR 27-JUL-1999; 99US-0145919.  
XX PR 28-JUL-1999; 99US-0145951.  
XX PR 02-AUG-1999; 99US-0146386.  
XX PR 02-AUG-1999; 99US-0146388.  
XX PR 02-AUG-1999; 99US-0146389.  
XX PR 03-AUG-1999; 99US-0147038.  
XX PR 04-AUG-1999; 99US-0147204.  
XX PR 04-AUG-1999; 99US-0147302.  
XX PR 05-AUG-1999; 99US-0147192.  
XX PR 05-AUG-1999; 99US-0147260.  
XX PR 06-AUG-1999; 99US-0147303.  
XX PR 06-AUG-1999; 99US-0147416.  
XX PR 09-AUG-1999; 99US-0147493.

```
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 23-SEP-1999; 99US-0155659.
PR 24-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 29-SEP-1999; 99US-0157117.
PR 04-OCT-1999; 99US-0157753.
PR 05-OCT-1999; 99US-0157865.
PR 06-OCT-1999; 99US-0158029.
PR 07-OCT-1999; 99US-0158232.
PR 08-OCT-1999; 99US-0158369.
PR 12-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      78.0%; Score 32; DB 21; Length 80;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Db |||||  
 6 pleely 11

Search completed: January 29, 2002, 10:21:52  
Job time: 428. sec



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 10:24:11 ; Search time 133.18 seconds  
(without alignments)  
1.352 Million cell updates/sec

Title: US-09-763-397A-21  
Perfect score: 41  
Sequence: 1 LTPLEELY 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles!.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	75.6	371	2	US-08-442-809A-76
2	30	73.2	1144	1	US-08-147-812-5
3	30	73.2	1144	2	US-08-319-866-12
4	30	73.2	1144	4	US-09-123-708-2
5	30	73.2	1144	4	US-09-123-624-2
6	30	73.2	1203	4	US-09-075-272-4
7	30	73.2	3169	2	US-08-477-451-6
8	29	70.7	102	3	US-08-946-329A-93
9	29	70.7	251	4	US-08-665-647-7
10	29	70.7	417	4	US-09-251-645-2
11	29	70.7	480	4	US-09-078-173A-25
12	29	70.7	480	4	US-09-537-357-32
13	29	70.7	519	4	US-09-312-183A-3
14	29	70.7	547	4	US-09-312-183A-2
15	29	70.7	1976	3	US-09-024-020B-9
16	29	70.7	1978	3	US-09-024-020B-3
17	29	70.7	1988	3	US-09-024-020B-4
18	28	68.3	9	1	US-08-139-054-7
19	28	68.3	11	1	US-08-139-054-5
20	28	68.3	11	1	US-08-139-054-9
21	28	68.3	14	1	US-08-139-054-6
22	28	68.3	142	4	US-08-945-983-7
23	28	68.3	204	1	US-08-292-945-8
24	28	68.3	204	1	US-08-252-073A-8
25	28	68.3	204	5	PCT-US93-1207A-8
26	28	68.3	340	2	US-08-578-592-5
27	28	68.3	340	3	US-09-185-111-5

28	28	68.3	342	4	US-09-461-474-16	Sequence 16, Appl
29	28	68.3	449	3	US-08-987-743-7	Sequence 7, Appl
30	28	68.3	462	2	US-08-477-451-24	Sequence 24, Appl
31	28	68.3	611	1	US-08-386-727-4	Sequence 4, Appl
32	28	68.3	611	2	US-08-600-452A-4	Sequence 4, Appl
33	28	68.3	863	2	US-08-666-271-2	Sequence 2, Appl
34	28	68.3	1079	2	US-08-929-967-8	Sequence 8, Appl
35	27	65.9	11	1	US-08-408-604A-32	Sequence 32, Appl
36	27	65.9	11	1	US-08-408-604A-43	Sequence 43, Appl
37	27	65.9	11	5	PCT-US93-09626-32	Sequence 32, Appl
38	27	65.9	11	5	PCT-US93-09626-43	Sequence 43, Appl
39	27	65.9	12	1	US-08-128-971B-11	Sequence 11, Appl
40	27	65.9	12	1	US-08-128-971B-14	Sequence 14, Appl
41	27	65.9	13	1	US-08-408-604A-36	Sequence 36, Appl
42	27	65.9	13	5	PCT-US93-09626-36	Sequence 36, Appl
43	27	65.9	18	4	US-08-652-877-39	Sequence 39, Appl
44	27	65.9	18	4	US-08-476-515A-39	Sequence 39, Appl
45	27	65.9	142	2	US-08-805-117-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1  
US-08-442-809A-76  
; Sequence 76, Application US/08442809A  
; Patent No. 5976873  
; GENERAL INFORMATION:  
; APPLICANT: Bohinski, Robert J.,  
; APPLICANT: Whitsett, Jeffrey A.  
; TITLE OF INVENTION: Nucleic Acid Sequences  
; TITLE OF INVENTION: Controlling Lung Cell -  
; TITLE OF INVENTION: Specific Gene Expression  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan,  
; STREET: 6 Becker Farm Road  
; CITY: Roseland  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch diskette  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WordPerfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/442,809A  
; FILING DATE: 17-MAY-1995  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/245,356  
; FILING DATE: 18-MAY-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Olstein, Elliot M.  
; REGISTRATION NUMBER: 24,025  
; REFERENCE/DOCKET NUMBER: 271010-360  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 76:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 371 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: human thyroid transcription factor-1  
; US-08-442-809A-76

us-09-763-397a-21.ra

Mon Feb 4 15:23:43 2002

TITLE OF INVENTION: ASSOCIATED WITH LONG-TERM MEMORY  
 NUMBER OF SEQUENCES: 24  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
 STREET: Two Militia Drive  
 CITY: Lexington  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02173

QY 1 LTPLEEL 8  
 Db 17 LSPLEEL 24

RESULT 2  
 US-08-147-812-5  
 ; Sequence 5, Application US/08147812  
 ; Patent No. 5766909  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Xie, Qiao-wen  
 ; APPLICANT: Nathan, Carl F.  
 ; APPLICANT: Mumford, Richard A.  
 ; APPLICANT: Calaycay, Jimmy Ramos  
 ; TITLE OF INVENTION: DNA Encoding Inducible Nitric Oxide Synthase  
 ; NUMBER OF SEQUENCES: 6  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Merck & Co., Inc.  
 ; STREET: 126 East Lincoln Avenue  
 ; CITY: Rahway  
 ; STATE: New Jersey  
 ; COUNTRY: USA  
 ; ZIP: 07065  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy Disk  
 ; COMPUTER: Macintosh Centris650  
 ; OPERATING SYSTEM: Macintosh 7.0.1  
 ; SOFTWARE: Microsoft Word 5.1a  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/147,812  
 ; FILING DATE: No. 5766909 Available  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/841,641  
 ; FILING DATE: 02-FEB-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Wallen, John W III  
 ; REGISTRATION NUMBER: 35,403  
 ; REFERENCE/DOCKET NUMBER: 186581A  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (908) 594-3905  
 ; TELEFAX: (908) 594-4720  
 ; TELEX: 138825  
 ; INFORMATION FOR SEQ ID NO: 5:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1144 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; US-08-147-812-5

Query Match 73.2%; Score 30; DB 1; Length 1144;  
 Best Local Similarity 100.0%; Pred. No. 6.6e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 TPLEEL 7  
 Db 128 TPLEEL 133

RESULT 3  
 US-08-319-866-12  
 ; Sequence 12, Application US/08319866  
 ; Patent No. 5923223  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tully, Timothy P.  
 ; APPLICANT: Yin, Jerry C.  
 ; APPLICANT: Regulski, Michael  
 ; TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF GENES

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/319,866  
 FILING DATE: 7-OCT-1994  
 CLASSIFICATION: 514  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Granahan, Patricia  
 REGISTRATION NUMBER: 32,227  
 REFERENCE/DOCKET NUMBER: CSHL94-03  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 861-6240  
 TELEFAX: (617) 861-9540  
 INFORMATION FOR SEQ ID NO: 12:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1144 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-319-866-12

Query Match 73.2%; Score 30; DB 2; Length 1144;  
 Best Local Similarity 100.0%; Pred. No. 6.6e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 TPLEEL 7  
 Db 128 TPLEEL 133

RESULT 4  
 US-09-123-708-2  
 ; Sequence 2, Application US/09123708  
 ; Patent No. 6146887  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Schraeder, Juergen  
 ; APPLICANT: GODECKE, Axel  
 ; TITLE OF INVENTION: DNA EXPRESSION VECTORS FOR USE IN GENE THERAPEUTIC  
 ; TREATMENT OF VASCULAR DISORDERS  
 ; FILE REFERENCE: 511169-2003  
 ; CURRENT APPLICATION NUMBER: US/09/123,708  
 ; CURRENT FILING DATE: 1998-07-28  
 ; EARLIER APPLICATION NUMBER: 08/553,503  
 ; EARLIER FILING DATE: 1996-03-01  
 ; EARLIER APPLICATION NUMBER: P4411402.8  
 ; EARLIER FILING DATE: 1994-03-31  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 2  
 ; LENGTH: 1144  
 ; TYPE: PRT  
 ; ORGANISM: Cytomegalovirus  
 ; US-09-123-708-2

Query Match 73.2%; Score 30; DB 4; Length 1144;

Best Local Similarity 100.0%; Pred. No. 6.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TPLEEL 7  
|||||  
Db 128 TPLEEL 133

## RESULT 5

US-09-123-624-2  
; Sequence 2, Application US/09123624  
; Patent No. 6149936  
; GENERAL INFORMATION:  
; APPLICANT: SCHRADER, Jurgen  
; APPLICANT: GODECKE, Axel  
; TITLE OF INVENTION: DNA EXPRESSION VECTORS FOR USE IN THE GENE THERAPEUTIC  
; TITLE OF INVENTION: TREATMENT OF VASCULAR DISORDERS  
; FILE REFERENCE: 511169-2004  
; CURRENT APPLICATION NUMBER: US/09/123,624  
; CURRENT FILING DATE: 1998-07-28  
; PRIOR APPLICATION NUMBER: 08/553,503  
; PRIOR FILING DATE: 1996-03-01  
; PRIOR APPLICATION NUMBER: 4411402.8  
; PRIOR FILING DATE: 1994-03-31  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 1144  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-123-624-2

Query Match 73.2%; Score 30; DB 4; Length 1144;  
Best Local Similarity 100.0%; Pred. No. 6.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TPLEEL 7  
|||||  
Db 128 TPLEEL 133

## RESULT 6

US-09-075-272-4  
; Sequence 4, Application US/09075272  
; Patent No. 6136598  
; GENERAL INFORMATION:  
; APPLICANT: MILLER, A. DUSTY  
; APPLICANT: WOLGAMOT, GREG  
; APPLICANT: BONHAM, LYNN  
; TITLE OF INVENTION: MUS DUNNI ENDOGENOUS RETROVIRAL  
; TITLE OF INVENTION: PACKAGING CELL LINES  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP  
; STREET: Two Embarcadero Center, 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 94111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/075,272  
; FILING DATE: 08-MAY-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/046,140  
; FILING DATE: 09-MAY-1997  
; ATTORNEY/AGENT INFORMATION:

; NAME: Poor, Brian W.  
; REGISTRATION NUMBER: 32,928  
; REFERENCE/DOCKET NUMBER: 14538A-003710  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 467-9600  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1203 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-075-272-4

Query Match 73.2%; Score 30; DB 4; Length 1203;  
Best Local Similarity 75.0%; Pred. No. 7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTPLEELY 8  
|||||  
Db 1035 LTPFEVLY 1042

## RESULT 7

US-08-477-451-6  
; Sequence 6, Application US/08477451  
; Patent No. 5928865  
; GENERAL INFORMATION:  
; APPLICANT: Covacci, Antonello  
; TITLE OF INVENTION: Helicobacter Pylori Cagi Region  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94608-2916  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/477,451  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mcclung, Barbara G.  
; REGISTRATION NUMBER: 33,113  
; REFERENCE/DOCKET NUMBER: 0335.002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 510-601-2708  
; TELEFAX: 510-655-3542  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3169 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-477-451-6

Query Match 73.2%; Score 30; DB 2; Length 3169;  
Best Local Similarity 85.7%; Pred. No. 2.1e+03;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTPLEEL 7  
|:|||||  
Db 2614 LSPLEEL 2620

us-09-763-397a-21.ra1

Mon Feb 4 15:23:43 2002

STREET: 2000 Pennsylvania Avenue, NW - Ste. 5500

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20006-1888

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

CURRENT APPLICATION DATA: Patent Release #1.0, Version #1.30

APPLICATION NUMBER: US/08/665,647

FILING DATE: 18-JUN-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Murashige, Kate H.

REGISTRATION NUMBER: 29,959

REFERENCE/DOCKET NUMBER: 22550-20025.25

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 887-1500

TELEFAX: (202) 822-0168

TELEX: 90-4030 MRSNFORSWH

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 251 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-665-647-7

Query Match 70.7%; Score 29; DB 2; Length 251;

Best Local Similarity 62.5%; Pred. No. 1.9e+02;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LTPLEELY 8

DB 150 LDIPIKELY 157

RESULT 10

US-09-251-645-2

Sequence 2, Application US/09251645

Patent No. 6281413

GENERAL INFORMATION:

APPLICANT: Kramer, Vance C.

APPLICANT: Morgan, Michael K.

APPLICANT: Anderson, Arne R.

APPLICANT: Hart, Hope

APPLICANT: Warren, Gregory W.

APPLICANT: Dunn, Martha

APPLICANT: Chen, Jeng S.

TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS

TITLE OF INVENTION: AND NUCLEIC ACID SEQUENCES CODING THEREFOR

FILE REFERENCE: CGC1963/A

CURRENT APPLICATION NUMBER: US/09/251,645

CURRENT FILING DATE: 1999-02-17

NUMBER OF SEQ ID NOS: 22

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 2

LENGTH: 417

TYPE: PRT

ORGANISM: Photorhabdus luminescens

US-09-251-645-2

Query Match 70.7%; Score 29; DB 4; Length 417;

Best Local Similarity 71.4%; Pred. No. 3.3e+02;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TPLEELY 8

DB 311 TPLNDLY 317

RESULT 8

US-08-946-329A-93

Sequence 93, Application US/08946329A

Patent No. 6057091

GENERAL INFORMATION:

APPLICANT: Beachy, Philip A.

APPLICANT: Porter, Jeffrey A.

TITLE OF INVENTION: NOVEL HEDGEHOG-DERIVED POLYPEPTIDES

NUMBER OF SEQUENCES: 109

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish &amp; Richardson P.C.

STREET: 4225 Executive Square, Suite 1400

CITY: La Jolla

STATE: CA

COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: FastSeq for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/946,329A

FILING DATE: 07-OCT-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/061,323

FILING DATE: 07-OCT-1996

APPLICATION NUMBER: 08/729,743

FILING DATE: 10-JUL-1996

APPLICATION NUMBER: 08/567,357

FILING DATE: 04-DEC-1995

APPLICATION NUMBER: 08/349,498

FILING DATE: 02-DEC-1994

ATTORNEY/AGENT INFORMATION:

NAME: Halle, Lisa A.

REGISTRATION NUMBER: 38,347

REFERENCE/DOCKET NUMBER: 07265/140001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619/678-5099

TELEFAX: 619/678-5099

INFORMATION FOR SEQ ID NO: 93:

SEQUENCE CHARACTERISTICS:

LENGTH: 102 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-946-329A-93

Query Match 70.7%; Score 29; DB 3; Length 102;

Best Local Similarity 62.5%; Pred. No. 69;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LTPLEELY 8

DB 18 LITMEEMY 25

RESULT 9

US-08-665-647-7

Sequence 7, Application US/08665647

Patent No. 5935803

GENERAL INFORMATION:

APPLICANT: Desqueux, Nicki J.

APPLICANT: Ron, Dorit

APPLICANT: Voronova, Anna F.

APPLICANT: Napolitano, Eugene W.

TITLE OF INVENTION: METHODS TO IDENTIFY IMMUNOMODULATORS

TITLE OF INVENTION: USING COGNATE INTERACTION OF PKC-THETA

NUMBER OF SEQUENCES: 89

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON &amp; FOERSTER

```

; FILE REFERENCE: 6617-13
; CURRENT APPLICATION NUMBER: US/09/312.183A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 519
; TYPE: PRT
; ORGANISM: Glycine max
US-09-312-183A-3

Query Match 70.7%; Score 29; DB 4; Length 519;
Best Local Similarity 62.5%; Pred. No. 4.3e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 LTPLEELY 8
| | | | |
Db 224 LTPIHETY 231

RESULT 14
US-09-312-183A-2
; Sequence 2, Application US/09312183A
; Patent No. 6303766
; GENERAL INFORMATION:
; APPLICANT: GRABAU, ELIZABETH A.
; APPLICANT: HEGEMAN, CARLA E.
; TITLE OF INVENTION: SOYBEAN PHYTASE AND NUCLEIC ACID ENCODING THE SAME
; FILE REFERENCE: 6617-13
; CURRENT APPLICATION NUMBER: US/09/312.183A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 547
; TYPE: PRT
; ORGANISM: Glycine max
US-09-312-183A-2

Query Match 70.7%; Score 29; DB 4; Length 547;
Best Local Similarity 62.5%; Pred. No. 4.5e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 LTPLEELY 8
| | | | |
Db 252 LTPIHETY 259

RESULT 15
US-09-024-020B-9
; Sequence 9, Application US/09024020B
; Patent No. 6030810
; GENERAL INFORMATION:
; APPLICANT: DELGADO, STEPHEN G.
; APPLICANT: DIETRICH, PAUL S.
; APPLICANT: FISH, LINDA M.
; APPLICANT: HERMAN, RONALD C.
; APPLICANT: SANGAMESWARAN, LAKSHMI
; TITLE OF INVENTION: NOVEL CLONED TETRODOTOXIN-SENSITIVE
; TITLE OF INVENTION: SODIUM CHANNEL I-SUBUNIT AND A SPLICE VARIANT THEREOF
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JANET PAULINE CLARK
; STREET: 3401 HILLVIEW AVENUE, MS A2-250
; CITY: PALO ALTO
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94304-1397
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
;

; FILE REFERENCE: 6617-13
; CURRENT APPLICATION NUMBER: US/09/078.173A
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 25
; LENGTH: 480
; TYPE: PRT
; ORGANISM: Capsicum annuum (green pepper)
US-09-078-173A-25

Query Match 70.7%; Score 29; DB 4; Length 480;
Best Local Similarity 62.5%; Pred. No. 3.9e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LTPLEELY 8
| | | | |
Db 229 LQPLEEIF 236

RESULT 12
US-09-537-357-32
; Sequence 32, Application US/09537357
; Patent No. 6271018
; GENERAL INFORMATION:
; APPLICANT: Alan Brash
; APPLICANT: Nathalie Tijet
; TITLE OF INVENTION: MUSKMELON (CUCUMIS MELO) HYDROPEROXIDE
; TITLE OF INVENTION: LYASE AND USES THEREOF
; FILE REFERENCE: 06027.0002
; CURRENT APPLICATION NUMBER: US/09/537.357
; CURRENT FILING DATE: 2000-03-29
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 480
; TYPE: PRT
; ORGANISM: Capsicum annuum (green pepper)
US-09-537-357-32

Query Match 70.7%; Score 29; DB 4; Length 480;
Best Local Similarity 62.5%; Pred. No. 3.9e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LTPLEELY 8
| | | | |
Db 229 LQPLEEIF 236

RESULT 13
US-09-312-183A-3
; Sequence 3, Application US/09312183A
; Patent No. 6303766
; GENERAL INFORMATION:
; APPLICANT: GRABAU, ELIZABETH A.
; APPLICANT: HEGEMAN, CARLA E.
; TITLE OF INVENTION: SOYBEAN PHYTASE AND NUCLEIC ACID ENCODING THE SAME
```

us-09-763-397a-21.ra1

Mon Feb 4 15:23:43 2002

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/024.020B
; FILING DATE: 16-FEB-1998
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/039,447
; FILING DATE: 26-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: CLARK, JANET P.
; REGISTRATION NUMBER: 34,799
; REFERENCE/DOCKET NUMBER: R0020B-REG
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 852-3097
; TELEFAX: (650) 855-5322
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1976 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-024-020B-9

```

```

Query Match          70.7%; Score 29; DB 3; Length 1976;
Best Local Similarity 83.3%; Pred. No. 1.9e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      3 PLEELY 8
        |||||
Db     1029 PLDELY 1034

```

```

Search completed: January 29, 2002, 10:24:12
Job time: 518 sec

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 10:26:47 ; Search time 144.96 Seconds  
(without alignments)  
4.204 Million cell updates/sec

Title: us-09-763-397a-21

Perfect score: 41

Sequence: 1 LTPLEELY 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	41	100.0	782	2 S27833	rhopty-associated
2	39	95.1	447	2 I37082	GDP-dissociation i
3	39	95.1	447	2 A35652	smg p25A regulator
4	39	95.1	447	2 B56024	GDP dissociation i
5	39	95.1	447	2 A54091	rab GDP dissociati
6	38	92.7	444	2 T27222	hypothetical prote
7	36	87.8	445	2 A56024	GDP dissociation i
8	36	87.8	445	2 C56956	GDP dissociation i
9	36	87.8	445	2 B54091	rab GDP dissociati
10	35	85.4	448	2 S36746	GDP dissociation i
11	34	82.9	335	2 G83207	hypothetical prote
12	34	82.9	390	2 E72411	phosphopentomutase
13	34	82.9	412	2 D81693	conserved hypothet
14	34	82.9	414	2 B71538	probable hypotheti
15	34	82.9	990	2 T03784	probable receptor
16	33	80.5	117	2 A69828	hypothetical prote
17	33	80.5	139	2 T26722	hypothetical prote
18	33	80.5	338	2 S74042	probable pyruvate
19	33	80.5	393	2 B83841	phosphopentomutase
20	33	80.5	394	2 B69619	phosphodiesteribomu
21	33	80.5	609	2 A64432	modulation factor
22	33	80.5	632	2 S64786	hypothetical prote
23	32	78.0	139	2 T02564	hypothetical prote
24	32	78.0	139	2 H84729	hypothetical prote
25	32	78.0	304	2 A32108	translation initia
26	32	78.0	338	2 S38030	suppressor protein
27	32	78.0	366	2 G85699	glucokinase (EC 2.
28	32	78.0	366	2 E64866	hypothetical prote
29	32	78.0	386	2 S37691	ran GTPase activat

30	32	78.0	393	2 JE0180	phosphopentomutase
31	32	78.0	411	2 H72084	cbs domain protein
32	32	78.0	411	2 H86539	CBS domain protein
33	32	78.0	435	2 S64950	protein kinase HOG
34	32	78.0	514	2 S46733	hypothetical prote
35	32	78.0	532	2 S76525	methionine--tRNA I
36	32	78.0	532	2 T27457	hypothetical prote
37	32	78.0	602	2 T13219	major capsid prote
38	32	78.0	630	1 BWUT8Q	regulatory protein
39	32	78.0	630	2 A36359	VSG expression sit
40	32	78.0	630	2 S13724	ESAG 8 protein - T
41	32	78.0	848	2 C70203	DNA topoisomerase
42	32	78.0	1296	2 S55511	valine--tRNA ligas
43	31	75.6	133	2 T35218	hypothetical prote
44	31	75.6	157	2 G71843	probable dGTP pyro
45	31	75.6	295	2 E84425	probable homeodoma

#### ALIGNMENTS

##### RESULT 1

S27833  
rhopty-associated protein 1 precursor - malaria parasite (Plasmodium falciparum)  
N:Alternate names: protective antigen  
C:Species: Plasmodium falciparum  
C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 09-Jun-2000  
C:Accession: A45514; S27833  
R:Ridley, R.G.; Takacs, B.; Lahm, H.W.; Delves, C.J.; Goman, M.; Certa, U.; Matile, H.  
Mol. Biochem. Parasitol. 41, 125-134, 1990  
A:Title: Characterisation and sequence of a protective rhopty antigen from Plasmodiu  
A:Reference number: A45514; MUID:90348711  
A:Accession: A45514  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-782 <R12>  
A:Cross-references: GB:M32853; NID:gl60656; PID:gl60657  
C:Superfamily: Plasmodium falciparum rhopty-associated protein 1

Query Match 100.0%; Score 41; DB 2; Length 782;  
Best Local Similarity 100.0%; Pred. No. 2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LTPLEELY 8  
Db 202 LTPLEELY 209

##### RESULT 2

I37082  
GDP-dissociation inhibitor XAP-4 - human  
C:Species: Homo sapiens (man)  
C:Date: 06-Sep-1996 #sequence\_revision 06-Sep-1996 #text\_change 21-Jul-2000  
C:Accession: I37082; I39293  
R:Sedlacek, Z.; Konecki, D.S.; Korn, B.; Klauck, S.M.; Poustka, A.  
Mamm. Genome 5, 633-639, 1994  
A:Title: Evolutionary conservation and genomic organization of XAP-4, an Xq28 located  
A:Reference number: I37082; MUID:95152170  
A:Accession: I37082  
A>Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-447 <RE>  
A:Cross-references: EMBL:X79354; NID:9695522; PIDN:CAA55909.1; PID:9695523  
A:Accession: I39293  
A>Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: mRNA  
A:Residues: 1-447 <RES>  
A:Cross-references: EMBL:X79353; NID:9695584; PID:9695585  
C:Genetics:  
A:Gene: GDB:RABGDIA  
A:Cross-references: GDB:I347097  
A:Introns: 15/3; 51/3; 85/1; 130/1; 196/2; 240/2; 273/3; 331/1; 379/2; 397/3

us-09-763-397a-21.rpr

Mon Feb 4 15:23:43 2002

C:Superfamily: human GDP dissociation inhibitor XAP-4

Query Match 95.1%; Score 39; DB 2; Length 447;  
Best Local Similarity 87.5%; Pred. No. 2.8;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTPLEELY 8  
:|||||  
Db 46 ITPLEELY 53

RESULT 3  
A35652  
smg p25A regulatory protein - bovine  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 28-Sep-1990 #sequence\_revision 20-Jun-2000  
C:Accession: A35652  
R:Matsui, Y.; Kikuchi, A.; Araki, S.; Hata, Y.; Kondo, J.; Teranishi, Y.; Takai, Y.  
Mol. Cell. Biol. 10, 4116-4122, 1990  
A:Title: Molecular cloning and characterization of a novel type of regulatory protein (GDP dissociation inhibitor) from rat liver cytosol of a GDP dissociation inhibitor XAP-4  
A:Reference number: A35652; MUID:90318376  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-447 <MAN>  
A:Cross-references: GB:D90103; NID:q217563; PIDN:BAAL4134.1; PID:q217564; GB:M55560  
C:Superfamily: human GDP dissociation inhibitor XAP-4

Query Match 95.1%; Score 39; DB 2; Length 447;  
Best Local Similarity 87.5%; Pred. No. 2.8;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTPLEELY 8  
:|||||  
Db 46 ITPLEELY 53

RESULT 4  
B56024  
GDP dissociation inhibitor 1 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 10-Nov-1995 #sequence\_revision 10-Nov-1995 #text\_change 26-May-2000  
C:Accession: B56024  
R:Shishava, A.; Suedhof, T.C.; Czech, M.P.  
Mol. Cell. Biol. 14, 3459-3468, 1994  
A:Title: Cloning, characterization, and expression of a novel GDP dissociation inhibitor from rat liver cytosol of a GDP dissociation inhibitor XAP-4  
A:Reference number: B56024; MUID:94217740  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-447 <SHI>  
A:Cross-references: GB:U07952; NID:g493530; PIDN:AB16909.1; PID:g516540  
C:Superfamily: human GDP dissociation inhibitor XAP-4

Query Match 95.1%; Score 39; DB 2; Length 447;  
Best Local Similarity 87.5%; Pred. No. 2.8;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTPLEELY 8  
:|||||  
Db 46 ITPLEELY 53

RESULT 5  
A54091  
rab GDP dissociation inhibitor alpha - rat  
N:Alternate names: rab GDI alpha  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 26-May-2000  
C:Accession: A54091; S35964; A37996

R:Nishimura, N.; Nakamura, H.; Takai, Y.; Sano, K.  
J. Biol. Chem. 269, 14191-14198, 1994  
A:Title: Molecular cloning and characterization of two rab GDI species from rat brain  
A:Reference number: A54091; MUID:94245743  
A:Accession: A54091  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-447 <NIS>  
A:Cross-references: GB:X74402  
R:Nishimura, N.  
submitted to the EMBL Data Library, July 1993  
A:Reference number: S35964  
A:Accession: S35964  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-255, 'D', 257-288, 'D', 290-447 <NI2>  
A:Cross-references: EMBL:X74402; NID:g396430; PIDN:CAAS2413.1; PID:g396431  
R:Ueda, T.; Takeyama, Y.; Ohmori, T.; Ohyanagi, H.; Saitoh, Y.; Takai, Y.  
Biochemistry 30, 909-917, 1991  
A:Title: Purification and characterization from rat liver cytosol of a GDP dissociation inhibitor XAP-4  
A:Reference number: A37996; MUID:91113701  
A:Accession: A37996  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 30-49; 104-112; 157-186; 270-285, 'T', 287-288, 'D', 290-293 <UED>  
C:Superfamily: human GDP dissociation inhibitor XAP-4

Query Match 95.1%; Score 39; DB 2; Length 447;  
Best Local Similarity 87.5%; Pred. No. 2.8;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTPLEELY 8  
:|||||  
Db 46 ITPLEELY 53

RESULT 6  
T27222  
hypothetical protein Y57G11C.10 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 26-May-2000  
C:Accession: T27222  
R:McMurray, A.  
submitted to the EMBL Data Library, September 1997  
A:Reference number: Z20330  
A:Accession: T27222  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-444 <WIL>  
A:Cross-references: EMBL:Z99281; PIDN:CAB16511.1; GSPDB:GN00022; CESP:Y57G11C.10  
A:Experimental source: clone Y57G11C  
C:Genetics:  
A:Gene: CESP:Y57G11C.10  
A:Map position: 4  
A:Introns: 332/1; 416/3  
C:Superfamily: human GDP dissociation inhibitor XAP-4

Query Match 92.7%; Score 38; DB 2; Length 444;  
Best Local Similarity 87.5%; Pred. No. 4.5;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTPLEELY 8  
:|||||  
Db 46 ITPLEELY 53

RESULT 7  
A56024  
GDP dissociation inhibitor 2 - mouse  
C:Species: Mus musculus (house mouse)



C>Date: 10-Nov-1995 #sequence\_revision 10-Nov-1995 #text\_change 26-May-2000

C:Accession: A56024

R:Shisheva, A.; Suedhof, T.C.; Czech, M.P.

Mol. Cell. Biol. 14, 3459-3468, 1994

A:Title: Cloning, characterization, and expression of a novel GDP dissociation inhibitor

A:Reference number: A56024; MUID:94217740

A:Accession: A56024

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-445 <SHI>

A:Cross-references: GB:U07951

C:Superfamily: human GDP dissociation inhibitor XAP-4

Query Match 87.8%; Score 36; DB 2; Length 445;

Best Local Similarity 75.0%; Pred. No. 12;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LTPLEELY 8

:|||||:

Db 46 ITPLEDLY 53

RESULT 8

C56956

GDP dissociation inhibitor beta - mouse

C:Species: Mus musculus (house mouse)

C>Date: 19-Oct-1995 #sequence\_revision 19-Oct-1995 #text\_change 26-May-2000

C:Accession: C56956

R:Janoueix-Lerosey, I.; Jollivet, P.; Camonis, J.; Marche, P.N.; Goud, B.

J. Biol. Chem. 270, 14801-14808, 1995

A:Title: Two-hybrid system screen with the small GTP-binding protein Rab6. Identification

A:Reference number: A56956; MUID:95301579

A:Accession: C56956

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-445 <JAN>

A:Cross-references: GB:L36314; NID:g538410; PIDN:AAA78786.1; PID:g538411

C:Superfamily: human GDP dissociation inhibitor XAP-4

Query Match 87.8%; Score 36; DB 2; Length 445;

Best Local Similarity 75.0%; Pred. No. 12;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LTPLEELY 8

:|||||:

Db 46 ITPLEDLY 53

RESULT 9

B54091

Rab GDP dissociation inhibitor beta - rat

N:Alternate names: rab GDI beta

C:Species: Rattus norvegicus (Norway rat)

C>Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 26-May-2000

C:Accession: B54091; S35965

R:Nishimura, N.; Nakamura, H.; Takai, Y.; Sano, K.

J. Biol. Chem. 269, 14191-14198, 1994

A:Title: Molecular cloning and characterization of two rab GDI species from rat brain: b

A:Reference number: A54091; MUID:94245743

A:Accession: B54091

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-445 <NIS>

A:Cross-references: GB:X74401; NID:g396432; PIDN:CAA52412.1; PID:g396433

R:Nishimura, N.

submitted to the EMBL Data Library, July 1993

A:Reference number: S35964

A:Accession: S35965

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-445 <NIS2>

A:Cross-references: EMBL:X74401; NID:g396432; PIDN:CAA52412.1; PID:g396433

C:Superfamily: human GDP dissociation inhibitor XAP-4

Query Match 87.8%; Score 36; DB 2; Length 445;

Best Local Similarity 75.0%; Pred. No. 12;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LTPLEELY 8

:|||||:

Db 46 ITPLEDLY 53

RESULT 10

S36746

GDP dissociation inhibitor - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C>Date: 09-Dec-1993 #sequence\_revision 12-Apr-1996 #text\_change 26-May-2000

C:Accession: S36746; S62097; S37806

R:Zahner, J.E.; Cheney, C.M.

Mol. Cell. Biol. 13, 217-227, 1993

A:Title: A Drosophila homolog of bovine smg p25a GDP dissociation inhibitor undergoes

A:Reference number: S36746; MUID:93109305

A:Accession: S36746

A:Molecule type: mRNA

A:Residues: 1-448 <ZAH1>

A:Cross-references: EMBL:L03209

A:Accession: S62097

A:Molecule type: protein

A:Residues: 24-47;83-106; 'L', 230-248;353-383 <ZAH3>

R:Zahner, J.E.; Cheney, C.M.

submitted to the EMBL Data Library, March 1993

A:Reference number: S37806

A:Accession: S37806

A:Molecule type: mRNA

A:Residues: 1-124; 'RPWHP', 130-448 <ZAH2>

A:Cross-references: EMBL:L03209; NID:g157491; PID:g157492

C:Genetics:

A:Gene: FlyBase:Gdi

A:Cross-references: FlyBase:FBgn0004868

C:Superfamily: human GDP dissociation inhibitor XAP-4

Query Match 85.4%; Score 35; DB 2; Length 448;

Best Local Similarity 75.0%; Pred. No. 19;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LTPLEELY 8

:|||||:

Db 47 ITPLEELF 54

RESULT 11

G83207

hypothetical protein PA3513 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: G83207

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: A82950; MUID:20437337

A:Accession: G83207

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-335 <STO>

A:Cross-references: GB:AE004771; GB:AE004091; NID:g9949650; PIDN:AAG06901.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA3513

```
Query Match      82.9%; Score 34; DB 2; Length 335;
Best Local Similarity 85.7%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 TPLEELY 8
Db 311 TPLEELY 317

RESULT 12
E72411
Phosphopentomutase - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: E72411
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey,
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 398, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316
A:Accession: E72411
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-390 <ARN>
A:Cross-references: GB:AE001701; GB:AE000512; NID:g4980648; PIDN:AAD35260.1; PID:g498066
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM0167
C:Superfamily: phosphopentomutase

Query Match      82.9%; Score 34; DB 2; Length 390;
Best Local Similarity 75.0%; Pred. No. 27;
Matches 6; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

QY 1 LTPLEELY 8
Db 163 IVPLEELY 170

RESULT 13
D81693
conserved hypothetical protein TC0527 [imported] - Chlamydia muridarum (strain Nigg)
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Jun-2000
C:Accession: D81693
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255
A:Accession: D81693
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-412 <TE>
A:Cross-references: GB:AE002321; GB:AE002160; NID:g7190560; PIDN:AAF39369.1; PID:g719056
A:Experimental source: strain Nigg (MoPn)
C:Genetics:
A:Gene: TC0527
C:Superfamily: hypothetical protein HI0107

Query Match      82.9%; Score 34; DB 2; Length 412;
Best Local Similarity 75.0%; Pred. No. 28;
Matches 6; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

QY 1 LTPLEELY 8
Db 213 LTPLEELY 220

RESULT 14
B71538
probable hypothetical protein containing cbs domains - Chlamydia trachomatis (serotyp
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 17-Mar-2000
C:Accession: B71538
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
A:Reference number: A71570; MUID:99000809
A:Accession: B71538
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-414 <ARN>
A:Cross-references: GB:AE001298; GB:AE001273; NID:g3328659; PIDN:AA67849.1; PID:g332
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: CT256
C:Superfamily: hypothetical protein HI0107

Query Match      82.9%; Score 34; DB 2; Length 414;
Best Local Similarity 75.0%; Pred. No. 28;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LTPLEELY 8
Db 213 LTPLEELY 220

RESULT 15
T03784
probable receptor protein kinase - rice
C:Species: Oryza sativa (rice)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000
C:Accession: T03784
R:Xu, Y.; Zhu, Q.; Lamb, C.
submitted to the EMBL Data Library, June 1995
A:Description: Molecular cloning and characterization of a rice gene encoding leucine
A:Reference number: Z15083
A:Accession: T03784
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-990 <XUY>
A:Cross-references: EMBL:X89226; NID:gl263159; PIDN:CAA61510.1
A:Experimental source: cv IR36, leaf
C:Genetics:
A:Gene: Irk2
A:Introns: 870/1
C:Superfamily: protein kinase Xa21; leucine-rich alpha-2-glycoprotein repeat homology

Query Match      82.9%; Score 34; DB 2; Length 990;
Best Local Similarity 75.0%; Pred. No. 75;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LTPLEELY 8
Db 217 LTPLEELY 224

Search completed: January 29, 2002, 10:26:47
Job time: 658 sec
```



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 11:13:48 ; Search time 80.65 seconds  
(without alignments)  
3.637 Million cell updates/sec

Title: US-09-763-397A-21  
Perfect score: 41  
Sequence: 1 LTPLEELY 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	39	95.1	447	1 GDIA_BOVIN	P21856 bos taurus
2	39	95.1	447	1 GDIA_CANFA	O97555 canis famil
3	39	95.1	447	1 GDIA_HUMAN	P31150 homo sapien
4	39	95.1	447	1 GDIA_RAT	P50398 rattus norv
5	36	87.8	445	1 GDIB_CANFA	O97556 canis famil
6	36	87.8	445	1 GDIB_HUMAN	P50395 homo sapien
7	36	87.8	445	1 GDIB_MOUSE	P50397 mus musculu
8	36	87.8	445	1 GDIC_MOUSE	Q61598 mus musculu
9	36	87.8	445	1 GDIC_RAT	P50399 rattus norv
10	33	80.5	396	1 DEOB_BAGSU	P46353 bacillus su
11	33	80.5	609	1 YA58_METJA	O58458 methanococ
12	32	78.0	304	1 IF2A_YEAST	P20459 saccharomyc
13	32	78.0	338	1 SD22_YEAST	P36047 saccharomyc
14	32	78.0	366	1 YCGT_ECOLI	P76015 escherichia
15	32	78.0	386	1 RNAL_SCHPO	P41391 schizosacch
16	32	78.0	424	1 MR15_DROME	O9Y011 drosophila
17	32	78.0	435	1 HOG1_YEAST	P32485 saccharomyc
18	32	78.0	514	1 YHK8_YEAST	P38776 saccharomyc
19	32	78.0	532	1 SYM_SYNY3	O55729 synecocyst
20	32	78.0	630	1 ESA8_TRYBB	P23799 trypanosoma
21	32	78.0	630	1 ESA8_TRYBO	P26337 trypanosoma
22	32	78.0	848	1 TOP1_BORBU	O51768 borrelia bu
23	31	75.6	157	1 NUDH_HELPJ	O92j28 helicobacte
24	31	75.6	371	1 TTF1_CANFA	P43698 canis famil
25	31	75.6	371	1 TTF1_HUMAN	P50220 mus musculu
26	31	75.6	372	1 TTF1_MOUSE	P50220 mus musculu
27	31	75.6	372	1 TTF1_RAT	P23441 rattus norv
28	31	75.6	578	1 PSP2_YEAST	P50109 saccharomyc
29	31	75.6	583	1 YCV1_YEAST	P25639 saccharomyc
30	31	75.6	1046	1 POL_FENV1	P31792 feline endo
31	31	75.6	2273	1 ABCR_HUMAN	P78363 homo sapien
32	30	73.2	105	1 GLN1_METBA	P54808 methanosarc
33	30	73.2	132	1 Y947_ARCFU	O29315 archaeoglob

RESULT 1					
GDIA_BOVIN					
ID GDIA_BOVIN	STANDARD:	PRT:	447 AA.		
AC P21856:					
DT 01-MAY-1991	(Rel. 18, Created)				
DT 01-MAY-1991	(Rel. 18, Last sequence update)				
DT 15-JUL-1998	(Rel. 36, Last annotation update)				
DE RAB GDP DISSOCIATION INHIBITOR ALPHA (RAB GDI ALPHA) (GDI-1) (SMG P25A					
DE GDI).					
GN GDI1 OR RABGDI1.					
OS Bos taurus (Bovine).					
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;					
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;					
OC Bovidae; Bovinae; Bos.					
OX NCBI_TaxID=9913;					
RN [1]					
RP SEQUENCE FROM N.A.					
RX MEDLINE=90318376; PubMed=2115118;					
RA Matsui Y., Kikuchi A., Araki S., Hata Y., Kondo J., Teranishi Y.,					
RA Takai Y.;					
RT "Molecular cloning and characterization of a novel type of regulatory					
RT protein (GDI) for smg p25A, a ras p21-like GTP-binding protein.";					
RL Mol. Cell. Biol. 10:4115-4122(1990).					
RN [2]					
RP SIMILARITY TO CHOROIDEAEMIA PROTEIN.					
RX MEDLINE=91270365; PubMed=1904992;					
RA Fodor E., Lee R.T., O'Donnell J.J.;					
RT "Analysis of choroideaemia gene.";					
RL Nature 351:614-614(1991).					
RN [3]					
RP X-RAY CRYSTALLOGRAPHY (1.81 ANGSTROMS).					
RX MEDLINE=96196507; PubMed=8609986;					
RA Schalk I., Zeng K., Wu S.-K., Stura E.A., Matteson J., Huang M.,					
RA Tandon A., Wilson I.A., Balch W.E.;					
RT "Structure and mutational analysis of Rab GDP-dissociation					
RT inhibitor.";					
RL Nature 381:42-48(1996).					
CC -1- FUNCTION: REGULATES THE GDP/GTP EXCHANGE REACTION OF MOST RAB					
CC PROTEINS BY INHIBITING THE DISSOCIATION OF GDP FROM THEM, AND THE					
CC SUBSEQUENT BINDING OF GTP TO THEM.					
CC -1- SIMILARITY: BELONGS TO THE TCD/MRS6 FAMILY OF GDP DISSOCIATION					
CC INHIBITOR.					
CC -----					
CC This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC the European Bioinformatics Institute. There are no restrictions on its					
CC use by non-profit institutions as long as its content is in no way					
CC modified and this statement is not removed. Usage by and for commercial					
CC entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>					
CC or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).					
CC -----					
DR EMBL; D90103; BAAL4134.1; -					
DR PIR; A35652; A35652.					
DR PDB; 1GND; 12-FEB-97.					
DR InterPro; IPR002005; Rab_GDI_REP.					
DR Pfam; PF00996; GDI; 1.					

P11987 methylococc  
O66417 aquifex aeo  
Q9zd28 rickettsia  
Q92207 candida alb  
O26043 helicobacte  
Q9zja7 helicobacte  
P98182 homo sapien  
P37198 homo sapien  
P77488 escherichia  
P42357 homo sapien  
P33492 mus musculu  
P21213 rattus norv

34 30 73.2 169 1 MEMG\_METCA  
35 30 73.2 193 1 Y226\_AQUAE  
36 30 73.2 368 1 Y165\_RICPR  
37 30 73.2 377 1 HOG1\_CANAL  
38 30 73.2 386 1 SELA\_HELPJ  
39 30 73.2 390 1 SELA\_HELPJ  
40 30 73.2 394 1 Z200\_HUMAN  
41 30 73.2 522 1 NU62\_HUMAN  
42 30 73.2 619 1 DXS\_ECOLI  
43 30 73.2 657 1 HUTH\_HUMAN  
44 30 73.2 657 1 HUTH\_MOUSE  
45 30 73.2 657 1 HUTH\_RAT

ALIGNMENTS

```
DR PRINTS; PRO0891; RABGDIREP.
DR PRINTS; PRO0892; RABGDI.
KW GTPase activation; 3D-structure.
SQ SEQUENCE 447 AA; 50565 MW; 69C68B03BDC1BA3F CRC64;

Query Match 95.1%; Score 39; DB 1; Length 447;
Best Local Similarity 87.5%; Pred. No. 0.82;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LTPLEELY 8
Db 46 ITPLEELY 53

RESULT 2
GDIA_CANFA STANDARD; PRT; 447 AA.
AC O97555;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RAB GDP DISSOCIATION INHIBITOR ALPHA (RAB GDI ALPHA) (GDI-1).
GN GD1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP MEDLINE=99019719; PubMed=9802909;
RA Chen W., Feng Y., Chen D., Wandinger-Ness A.;
RT "Rab11 is required for trans-golgi network-to-plasma membrane
RL Mol. Biol. Cell 9:3241-3257(1998).
CC -1- FUNCTION: REGULATES THE GDP/GTP EXCHANGE REACTION OF MOST RAB
CC PROTEINS BY INHIBITING THE DISSOCIATION OF GDP FROM THEM, AND THE
CC SUBSEQUENT BINDING OF GTP TO THEM.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE TCD/MRS6 FAMILY OF GDP DISSOCIATION
CC INHIBITOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF027360; AAD04246.1; -.
DR HSP; P21856; 1GND.
DR InterPro; IPR002005; Rab_GDI_REP.
DR Pfam; PF00996; GDI_1.
DR PRINTS; PRO0891; RABGDIREP.
KW GTPase activation.
SQ SEQUENCE 447 AA; 50520 MW; 10280DAD33E4BCD0 CRC64;

Query Match 95.1%; Score 39; DB 1; Length 447;
Best Local Similarity 87.5%; Pred. No. 0.82;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LTPLEELY 8
Db 46 ITPLEELY 53

RESULT 3
GDIA_HUMAN STANDARD; PRT; 447 AA.
ID GDIA_HUMAN
AC P31150; P50394;
DT 01-JUL-1993 (Rel. 26, Created)
```

```
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RAB GDP DISSOCIATION INHIBITOR ALPHA (RAB GDI ALPHA) (GDI-1) (XAP-4).
GN GD11 OR RABGDIA OR XAP4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=95152170; PubMed=7849400;
RA Sedlacek Z., Konecki D.S., Korn B., Kluack S.M., Poustka A.;
RT "Evolutionary conservation and genomic organization of XAP-4, an Xq28
RT located gene coding for a human rab GDP-dissociation inhibitor
RT (GDI).";
RL Mamm. Genome 5:633-639(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96062207; PubMed=7595614;
RA Nishimura N., Goji J., Nakamura H., Orita S., Takai Y., Sano K.;
RT "Cloning of a brain-type isoform of human Rab GDI and its expression
RT in human neuroblastoma cell lines and tumor specimens.";
RL Cancer Res. 55:5445-5450(1995).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=96311563; PubMed=8733135;
RA Chen E.Y., Zollo M., Mazzarella R.A., Ciccodicola A., Chen C.-N.,
RA Zuo L., Heiner C., Burrough F.W., Ripetto M., Schlessinger D.,
RA D'Urso M.;
RT "Long-range sequence analysis in Xq28: thirteen known and six
RT candidate genes in 219.4 kb of high GC DNA between the RCP/GCP and
RT G6PD loci.";
RL Hum. Mol. Genet. 5:659-668(1996).
RN [4]
RP SEQUENCE OF 143-181 FROM N.A.
RA Hochgeschwender U.;
RT Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 328-436 FROM N.A.
RA Bhat K.S.;
RT Submitted (XXY-1992) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 349-361.
RC TISSUE=Keratinocytes;
RX MEDLINE=93162043; PubMed=1286667;
RA Rasmussen H.H., van Damme J., Puype M., Gesser B., Celis J.E.,
RA Vandekerckhove J.;
RT "Microsequences of 145 proteins recorded in the two-dimensional gel
RT protein database of normal human epidermal keratinocytes.";
RL Electrophoresis 13:960-969(1992).
RN [7]
RP TISSUE SPECIFICITY.
RX MEDLINE=95359978; PubMed=7543319;
RA Bachner D., Sedlacek Z., Korn B., Hameister H., Poustka A.;
RT "Expression patterns of two human genes coding for different rab GDP-
RT dissociation inhibitors (GDIs), extremely conserved proteins involved
RT in cellular transport.";
RL Hum. Mol. Genet. 4:701-708(1995).
RN [8]
RP VARIANT XLMR PRO-92.
RX MEDLINE=98282090; PubMed=9620768;
RA D'Adamo P., Menegon A., Lo Nigro C., Grasso M., Gulisano M.,
RA Tamanini F., Bienvenu T., Gedeon A.K., Oostra B., Wu S.-K., Tandon A.,
RA Valtorta F., Balch W.E., Chelly J., Toniolo D.;
RT "Mutations in GD11 are responsible for x-linked non-specific mental
RT retardation.";
RL Nat. Genet. 19:134-139(1998).
RN [9]
RP VARIANT XLMR PRO-423.
RX MEDLINE=98334551; PubMed=9668174;
RA Bienvenu T., Des Portes V., Saint Martin A., McDonnell N., Billuart P.,
RA Carrie A., Vinet M.-C., Couvert P., Toniolo D., Ropers H.-H.,
RA
```



Mon Feb 4 15:23:43 2002

```

QY 1 LTPLEELY 8
DB 46 ITPLEELY 53

RESULT 5
GDIB_CANFA STANDARD; PRT; 445 AA.
AC 097556;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RAB GDP DISSOCIATION INHIBITOR BETA (RAB GDI BETA) (GDI-2).
GN GDI2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=9019719; PubMed=9802909;
RX Chen W., Feng Y., Chen D., Wandinger-Ness A.;
RT "Rab11 is required for trans-golgi network-to-plasma membrane
RT transport and a preferential target for GDP dissociation inhibitor.";
CC Mol. Biol. Cell 9:3241-3257(1998).
CC -1- FUNCTION: REGULATES THE GDP/GTP EXCHANGE REACTION OF MOST RAB
CC PROTEINS BY INHIBITING THE DISSOCIATION OF GDP FROM THEM, AND THE
CC SUBSEQUENT BINDING OF GTP TO THEM (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE TCD/MRS6 FAMILY OF GDP DISSOCIATION
CC INHIBITOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF027361; AAD04247.1; -.
CC DR HSSP; P21856; LGND.
CC DR InterPro; IPR002005; Rab_GDI_REP.
CC DR Pfam; PF00996; GDI; 1.
CC DR PRINTS; PR00891; RABGDIREP.
CC DR GTPase activation.
CC KW GTPase activation.
CC SQ SEQUENCE 445 AA; 50321 MW; 9CE6B5407A3B0400 CRC64;

Query Match 87.8%; Score 36; DB 1; Length 445;
Best Local Similarity 75.0%; Pred. No. 3.7;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTPLEELY 8
DB 46 ITPLEELY 53

RESULT 6
GDIB_HUMAN STANDARD; PRT; 445 AA.
AC P50395; O43928; Q90QM6;
DT 01-OCT-1996 (Rel. 34, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RAB GDP DISSOCIATION INHIBITOR BETA (RAB GDI BETA) (GDI-2).
GN GDI2 OR RABGDI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Brain;
AS Asada M., Kaibuchi K., Takai Y.;
Submitted (APR-1993) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
RA MEDLINE=98096592; PubMed=9434952;
RX Sedlacek Z., Muntermann E., Mincheva A., Lichter P., Poutska A.;
RT "The human rab GDI beta gene with long retroposon-rich introns maps to
RT 10p15 and its pseudogene to 7p11-p13.";
RL Mamm. Genome 9:78-80(1998).
RN [3]
RP SEQUENCE OF 81-439 FROM N.A.
RA TISSUE=Pancreas;
RX MEDLINE=20453283; PubMed=10996854;
RA Caillol N., Pasqualini E., Llobes R., Lombardo D.;
RT "Impairment of bile salt-dependent lipase secretion in human
RT pancreatic tumoral SOJ-6 cells.";
RL J. Cell. Biochem. 79:628-647(2000).
RN [4]
RP TISSUE SPECIFICITY.
RX MEDLINE=95359978; PubMed=7543319;
RA Bachner D., Sedlacek Z., Korn B., Hameister H., Poustka A.;
RT "Expression patterns of two human genes coding for different rab GDP-
RT dissociation inhibitors (GDIs), extremely conserved proteins involved
RT in cellular transport.";
RL Hum. Mol. Genet. 4:701-708(1995).
CC -1- FUNCTION: REGULATES THE GDP/GTP EXCHANGE REACTION OF MOST RAB
CC PROTEINS BY INHIBITING THE DISSOCIATION OF GDP FROM THEM, AND THE
CC SUBSEQUENT BINDING OF GTP TO THEM.
CC -1- TISSUE SPECIFICITY: UBIQUITOUS.
CC -1- SIMILARITY: BELONGS TO THE TCD/MRS6 FAMILY OF GDP DISSOCIATION
CC INHIBITOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D13988; BAA03095.1; -.
CC DR EMBL; Y13286; CAA73734.1; -.
CC DR EMBL; Y13287; CAA73735.1; -.
CC DR EMBL; Y13288; CAA73735.1; JOINED.
CC DR EMBL; Y13289; CAA73735.1; JOINED.
CC DR EMBL; Y13290; CAA73735.1; JOINED.
CC DR EMBL; Y13291; CAA73735.1; JOINED.
CC DR EMBL; Y13292; CAA73735.1; JOINED.
CC DR EMBL; Y13293; CAA73735.1; JOINED.
CC DR EMBL; Y13294; CAA73735.1; JOINED.
CC DR EMBL; Y13295; CAA73735.1; JOINED.
CC DR EMBL; Y13296; CAA73735.1; JOINED.
CC DR EMBL; Y13297; CAA73735.1; JOINED.
CC DR EMBL; AF144713; AAD34588.1; -.
CC DR HSSP; P21856; LGND.
CC DR MIM; 600767; -.
CC DR InterPro; IPR002005; Rab_GDI_REP.
CC DR Pfam; PF00996; GDI; 1.
CC DR PRINTS; PR00891; RABGDIREP.
CC DR PRINTS; PR00892; RABGDI.
CC KW GTPase activation.
CC FT CONFLICT 2 2 N -> D (IN REF. 1).
CC SQ SEQUENCE 445 AA; 50663 MW; CE186A2E3A47FCC9 CRC64;

Query Match 87.8%; Score 36; DB 1; Length 445;
Best Local Similarity 75.0%; Pred. No. 3.7;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTPLEELY 8
DB 46 ITPLEELY 53
```



```

RESULT 7
GDIB_MOUSE
ID GDIB_MOUSE STANDARD; PRT; 445 AA.
AC P50397;
DT 01-OCT-1996 (Rel. 34, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RAB GDP DISSOCIATION INHIBITOR BETA (RAB GDI BETA) (GDI-2).
CN GDI2 OR RABGDI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Skeletal muscle;
RX MEDLINE=94217740; PubMed=7513052;
RA Shisheva A., Suedhof T.C., Czech M.P.;
RT "Cloning, characterization, and expression of a novel GDP
RT dissociation inhibitor isoform from skeletal muscle.";
RL Mol. Cell. Biol. 14:3459-3468(1994).
CC -1- FUNCTION: REGULATES THE GDP/GTP EXCHANGE REACTION OF MOST RAB
CC PROTEINS BY INHIBITING THE DISSOCIATION OF GDP FROM THEM, AND THE
CC SUBSEQUENT BINDING OF GTP TO THEM.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC AND MEMBRANE-ASSOCIATED.
CC -1- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED.
CC -1- SIMILARITY: BELONGS TO THE TCD/MRS6 FAMILY OF GDP DISSOCIATION
CC INHIBITOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U07951; AAB16908.1; -.
CC DR HSSP; P21856; LGND.
CC DR MGD; MGI:99845; Gdi2.
CC DR InterPro; IPR002005; Rab_GDI_REP.
CC DR Pfam; PF00996; GDI; 1.
CC DR PRINTS; PR00891; RABGDIREP.
CC DR PRINTS; PR00892; RABGDI.
CC KW GTPase activation.
CC SQ SEQUENCE 445 AA; 50512 MW; 7FFD92EAE950EE49 CRC64;

Query Match 87.8%; Score 36; DB 1; Length 445;
Best Local Similarity 75.0%; Pred. No. 3.7;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LTPLEELY 8
Db :|||||
46 ITPLEDLY 53

RESULT 8
GDIC_MOUSE
ID GDIC_MOUSE STANDARD; PRT; 445 AA.
AC Q61598; Q9D8M9;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RAB GDP DISSOCIATION INHIBITOR BETA-2 (RAB GDI BETA-2) (GDI-3).
CN GDI3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.

```

```

RX MEDLINE=95301579; PubMed=7782346;
RA Janoueix-Lerosey I., Jollivet F., Camonis J., Marche P.N., Goud B.;
RT "Two-hybrid system screen with the small GTP-binding protein Rab6.
RT Identification of a novel mouse GDP dissociation inhibitor isoform and
RT two other potential partners of Rab6.";
RL J. Biol. Chem. 270:14801-14808(1995).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Pancreas;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -1- FUNCTION: REGULATES THE GDP/GTP EXCHANGE REACTION OF MOST RAB
CC PROTEINS BY INHIBITING THE DISSOCIATION OF GDP FROM THEM, AND THE
CC SUBSEQUENT BINDING OF GTP TO THEM.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC AND MEMBRANE-ASSOCIATED (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE TCD/MRS6 FAMILY OF GDP DISSOCIATION
CC INHIBITOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L36314; AAA78786.1; -.
CC DR EMBL; AK007871; BAB25321.1; -.
CC DR HSSP; P21856; LGND.
CC DR MGD; MGI:105939; Gdi3.
CC DR InterPro; IPR002005; Rab_GDI_REP.
CC DR Pfam; PF00996; GDI; 1.
CC DR PRINTS; PR00891; RABGDIREP.
CC KW GTPase activation.
CC FT CONFLICT 298 I -> N (IN REF. 2).
CC SQ SEQUENCE 445 AA; 50537 MW; 483B7C7456B30525 CRC64;

Query Match 87.8%; Score 36; DB 1; Length 445;
Best Local Similarity 75.0%; Pred. No. 3.7;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LTPLEELY 8
Db :|||||
46 ITPLEDLY 53

RESULT 9
GDIC_RAT
ID GDIC_RAT STANDARD; PRT; 445 AA.
AC P50399;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)

```

Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.

```

DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE RAB GDP DISSOCIATION INHIBITOR BETA-2 (RAB GDI BETA-2) (GDI-3).
GN GDI3.
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
ON [1]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RA Nishimura N., Nakamura H., Takai Y., Sano K.;
RT "Molecular cloning and characterization of two rab GDI species from
RL rat brain: brain-specific and ubiquitous types.";
RL J. Biol. Chem. 269:14191-14198(1994).
RN [2]
RP SEQUENCE OF 30-54 AND 58-74, AND CHARACTERIZATION.
RA MEDLINE=95298038; PubMed=7779099;
RC Ataki K., Nakanishi H., Hirano H., Kato M., Sasaki T., Takai Y.;
RT "Purification and characterization of Rab GDI beta from rat brain.";
RL Biochem. Biophys. Res. Commun. 211:296-305(1995).
CC -1- FUNCTION: REGULATES THE GDP/GTP EXCHANGE REACTION OF MOST RAB
CC PROTEINS BY INHIBITING THE DISSOCIATION OF GDP FROM THEM, AND THE
CC SUBSEQUENT BINDING OF GTP TO THEM.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC AND MEMBRANE-ASSOCIATED.
CC -1- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED.
CC -1- SIMILARITY: BELONGS TO THE TCD/MRS6 FAMILY OF GDP DISSOCIATION
CC INHIBITOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X74401; CAA52412.1; -
CC HSP: P21856; IGND.
CC InterPro: IPR002005; Rab_GDI_REP.
CC Pfam: PF00996; GDI; 1.
CC PRINTS: PR00891; RABGDIREP.
CC PRINTS: PR00892; RABGDI.
CC GTPase activation.
CC SEQUENCE 445 AA; 50685 MW; D081ABB31DB9CD29 CRC64;
SQ
Query Match 87.8%; Score 36; DB 1; Length 445;
Best Local Similarity 75.0%; Pred. No. 3.7; Mismatches 0; Indels 0; Gaps 0;
Matches 6; Conservative 2;
QY 1 LTPLEELY 8
DB 46 ITPLEEDLY 53
RESULT 10
DEOB_BACSU STANDARD; PRT; 396 AA.
AC P46353;
AT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PHOSPHOPENTOMUTASE (EC 5.4.2.7) (PHOSPHOEXORIBOMUTASE).
GN DRM.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
ON [1]
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BR151;
RA Schuch R., Piggot P.J., Garibian A., Nygaard P.;

```

---

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RA Kobayashi Y., Mizuno M., Masuda S., Takemaru K., Hosono S.,
RA Sato T., Takeuchi M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: D-RIBOSE 1-PHOSPHATE = D-RIBOSE 5-PHOSPHATE.
CC -1- CATALYTIC ACTIVITY: 2-DEOXY-D-RIBOSE 1-PHOSPHATE = 2-DEOXY-
CC D-RIBOSE 5-PHOSPHATE.
CC -1- PATHWAY: INVOLVED IN PURINE NUCLEOSIDE SALVAGE.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOPENTOMUTASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U32685; AAA74433.1; -
CC EMBL: D84432; BAA12650.1; -
CC EMBL: Z99116; CAB14282.1; -
CC Subtilisin; BG11331; drn.
CC InterPro: IPR002599; Metalloenzyme.
CC Pfam: PF01676; Metalloenzyme; 1.
CC Isoenzyme: Complete proteome.
CC CONFLICT 78 78 M -> L (IN REF. 2).
CC CONFLICT 209 210 MISSING (IN REF. 2).
CC CONFLICT 318 318 E -> G (IN REF. 2).
CC SEQUENCE 396 AA; 44294 MW; C41A3BE4317456E CRC64;
SQ
Query Match 80.5%; Score 33; DB 1; Length 396;
Best Local Similarity 75.0%; Pred. No. 14; Mismatches 1; Indels 0; Gaps 0;
Matches 6; Conservative 1;
QY 1 LTPLEELY 8
DB 169 VVPLEELY 176
RESULT 11
YA58_METJA STANDARD; PRT; 609 AA.
AC Q58458;
AT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN MJ1058.
GN MJ1058.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
ON [1]
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2651 / ATCC 43067;
RA MEDLINE=96337999; PubMed=8688087;
RA Sutton C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Bult C.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadov P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RA jannaschii.";
RL Science 273:1058-1073(1996).
CC -1- SIMILARITY: BELONGS TO THE NODU / CMCH FAMILY.
CC

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL: U67549; AAB99062.1; -  
DR TIGR: MJ1058; -  
DR InterPro: IPR003696; Carbamoyltransferase.  
DR Pfam: PF02543; CmcH\_NodU; 1.  
KW Hypothetical protein; Transferase; Complete proteome.  
SQ SEQUENCE 609 AA; 71733 MW; C91899CF01CB84EA CRC64;

Query Match 80.58; Score 33; DB 1; Length 609;  
Best Local Similarity 85.78; Pred. No. 23;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TPLEELY 8  
|||  
Db 359 TPFEELY 365

RESULT 12  
IF2A\_YEAST STANDARD; PRT; 304 AA.  
AC P20459;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE EUKARYOTIC TRANSLATION INITIATION FACTOR 2 ALPHA SUBUNIT (EIF-2-  
DE ALPHA).  
GN SU12 OR TIF211 OR YJR007W OR J1429.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89202411; PubMed=2649894;  
RA Cigan A.M., Pabich E.K., Feng L., Donahue T.F.;  
RT "Yeast translation initiation suppressor su12 encodes the alpha  
RT subunit of eukaryotic initiation factor 2 and shares sequence  
RT identity with the human alpha subunit".  
RL Proc. Natl. Acad. Sci. U.S.A. 86:2784-2788(1989).  
RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / FY1679;  
RA de Haan M., Smits P.H.M., Grivell L.A.;  
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: EIF-2 FUNCTIONS IN THE EARLY STEPS OF PROTEIN SYNTHESIS  
CC BY FORMING A TERNARY COMPLEX WITH GTP AND INITIATOR TRNA. THIS  
CC COMPLEX BINDS TO A 40S RIBOSOMAL SUBUNIT, FOLLOWED BY MRNA BINDING  
CC TO FORM A 43S PREINITIATION COMPLEX. JUNCTION OF THE 60S RIBOSOMAL  
CC SUBUNIT TO FORM THE 80S INITIATION COMPLEX IS PRECEDED BY  
CC HYDROLYSIS OF THE GTP BOUND TO EIF-2 AND RELEASE OF AN EIF-2-GDP  
CC BINARY COMPLEX. IN ORDER FOR EIF-2 TO RECYCLE AND CATALYZE ANOTHER  
CC ROUND OF INITIATION, THE GDP BOUND TO EIF-2 MUST EXCHANGE WITH GTP  
CC BY WAY OF A REACTION CATALYZED BY EIF-2B.  
CC -1- SUBUNIT: HETEROTRIMER COMPOSED OF AN ALPHA, A BETA, AND A GAMMA  
CC CHAIN.

CC -1- SIMILARITY: BELONGS TO THE EIF-2-ALPHA FAMILY.  
CC -1- SIMILARITY: CONTAINS 1 'SI MOTIF' DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC EMBL: M25552; AAA70332.1; -  
DR EMBL: X87611; CAA60929.1; -  
DR EMBL: Z49507; CAA89529.1; -  
DR PIR: A32108; A32108.  
DR SGD: S0003767; SU12.  
DR InterPro: IPR003029; S1.  
DR Pfam: PF00575; S1; 1.  
DR SMART: SM00316; S1; 1.  
KW Initiation factor; Protein biosynthesis; RNA-binding;  
KW Phosphorylation.  
FT MOD\_RES 52 PHOSPHORYLATION (BY GCN2).  
SQ SEQUENCE 304 AA; 34717 MW; AF4FIC80303A4E98 CRC64;

Query Match 78.08; Score 32; DB 1; Length 304;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PLEELY 8  
|||||  
Db 122 PLEELY 127

RESULT 13  
SD22\_YEAST STANDARD; PRT; 338 AA.  
AC P36047;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE PROTEIN PHOSPHATASES PP1 REGULATORY SUBUNIT SDS22.  
GN SDS22 OR EGP1 OR YKL193C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=YN 295;  
RA Mackelvie S.H., Andrews P.D., Stark M.J.R.;  
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Maia E Silva A., Bossier P., Vilela C., Fernandes L., Soares H.,  
RA Guerreiro P., Rodrigues-Pousada C.;  
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
RN [3]

RP SEQUENCE OF 1-136 FROM N.A.  
RC STRAIN=GRF88;  
RX MEDLINE=93348778; PubMed=8394042;  
RA Cheret G., Matheakis L.C., Sor F.;  
RT "DNA sequence analysis of the YCN2 region of chromosome XI in  
RT Saccharomyces cerevisiae".  
RL Yeast 9:661-667(1993).  
CC -1- FUNCTION: POTENTIAL REGULATOR OF THE MITOTIC FUNCTION OF YEAST  
CC TYPE 1 PROTEIN PHOSPHATASE.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -1- SIMILARITY: CONTAINS 8 LEUCINE-RICH REPEATS (LRR).  
CC -1- SIMILARITY: STRONG, TO S.POMBE SDS22 AND C.ELEGANS T09A5.9.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X83609; CAA58588.1; -  
DR EMBL: Z28193; CAA82037.1; -  
DR EMBL: X69765; -; NOT\_ANNOTATED\_CDS.  
DR PIR: S38030; S38030.

```
DR SCD; S0001676; SDS22.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR003592; LRR_out.
DR Pfam: PF00560; LRR; 9.
DR PRINTS: PR00019; LEURICHRPT.
DR SMART: SM00370; LRR; 11.
KW Leucine-rich repeat; Repeat; Mitosis; Nuclear protein.
FT REPEAT 65 89
FT REPEAT 91 112
FT REPEAT 112 136
FT REPEAT 156 179
FT REPEAT 181 202
FT REPEAT 202 226
FT REPEAT 246 270
FT REPEAT 295 318
SQ SEQUENCE 338 AA; 38887 MW; 1EF434C7276ABAE CRC64;

Query Match 78.0%; Score 32; DB 1; Length 338;
Best Local Similarity 87.5%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LTPLEELY 8
DB 224 LTNLEELY 231

RESULT 14
YCGT_ECOLI
ID YCGT_ECOLI STANDARD; PRT; 366 AA.
AC P76015;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 39.5 KDA PROTEIN IN TREA-PTH INTERGENIC REGION.
GN YCGT OR B1200.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampaio G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RA "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
CC -1- SIMILARITY: BELONGS TO THE DIHYDROXYACETONE KINASE FAMILY.
CC CORRESPONDS TO THE N-TERMINAL PART OF DHAK. YCGS IS THE C-TERMINAL
CC PART.
```

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000218; AAC74284.1; -.
DR EMBL; D90754; BAA36057.1; -.
DR EcoGene; EGI3901; ycgT.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 366 AA; 39494 MW; 08EA76FC216E55FC CRC64;

Query Match 78.0%; Score 32; DB 1; Length 366;
Best Local Similarity 85.7%; Pred. No. 22;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TPLEELY 8
DB 298 TPLSELY 304

RESULT 15
RNAL_SCHPO
ID RNAL_SCHPO STANDARD; PRT; 386 AA.
AC P41391;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RAN GPASE ACTIVATING PROTEIN 1 (PROTEIN RNAL).
GN RNAL OR SPAC22E12.07.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX STRAIN=972;
RX MEDLINE=93385646; PubMed=83741168;
RA Melchior F., Weber K., Gerke V.;
RA "A functional homologue of the RNAL gene product in
RT Schizosaccharomyces pombe: Purification, biochemical
RT characterization, and identification of a leucine-rich repeat
RT motif.";
RL Mol. Biol. Cell 4:569-581(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=972;
RX MEDLINE=93385646; PubMed=83741168;
RA Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RA Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CHARACTERIZATION, AND PARTIAL SEQUENCE.
RX STRAIN=972;
RX MEDLINE=95183543; PubMed=7878053;
RA Bischoff F.R., Krebber H., Kempf T., Hermes L., Ponstingl H.;
RA "Human RangPase-activating protein RangAP1 is a homologue of yeast
RT RnaIp involved in mRNA processing and transport.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:1749-1753(1995).
RN [4]
RX X-RAY CRYSTALLOGRAPHY (2.66 ANGSTROMS).
RX MEDLINE=99322670; PubMed=10394366;
RA Hillig R.C., Renault L., Vetter I.R., Drell T. IV, Wittinghofer A.,
RA Becker J.;
RA "The crystal structure of rnaIp: a new fold for a GTPase-activating
RT protein.";
RL Mol. Cell 3:781-791(1999).
CC -1- FUNCTION: GTPASE ACTIVATOR FOR THE NUCLEAR RAS-RELATED REGULATORY
CC PROTEIN SP1 (RAN), CONVERTING IT TO THE PUTATIVELY INACTIVE GDP-
CC BOUND STATE.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; POSSIBLY ENRICHED IN THE
CC NUCLEAR PERIPHERY.
CC -1- SIMILARITY: CONTAINS 8 LEUCINE-RICH REPEATS (LRR).
CC -1- SIMILARITY: TO MAMMALIAN RANGAP1.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X69882; CAA49509.1; -;  
DR EMBL; 270043; CAA93894.1; -;  
DR PIR; S37691; S37691.  
DR PDB; 1YRG; 29-MAR-00.  
DR InterPro; IPR001611; LRR.  
DR InterPro; IPR003592; LRR\_out.  
DR Pfam; PF00560; LRR; 2.  
DR SMART; SM00370; LRR; 5.  
KW GTPase activation; Repeat; Leucine-rich repeat; 3D-structure.  
FT REPEAT 23 48 LRR 1.  
FT REPEAT 85 112 LRR 2.  
FT REPEAT 113 141 LRR 3.  
FT REPEAT 179 206 LRR 4.  
FT REPEAT 207 235 LRR 5.  
FT REPEAT 236 264 LRR 6.  
FT REPEAT 265 293 LRR 7.  
FT REPEAT 294 322 LRR 8.  
FT DOMAIN 323 374 ASP/GLU-RICH (ACIDIC).  
FT SITE 74 74 CRITICAL RESIDUE FOR GTP HYDROLYSIS.  
SQ SEQUENCE 386 AA; 43235 MW; 8E1C9C506988A7F1 CRC64;

Query Match 78.08; Score 32; DB 1; Length 386;  
Best Local Similarity 85.7%; Pred. No. 23;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 TPLEELY 8  
    |||||  
Db 122 TPLEHLY 128

Search completed: January 29, 2002, 11:13:49  
Job time: 825 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 11:12:15 ; Search time 285.36 Seconds  
(without alignments)  
4.101 Million cell updates/sec

Title: US-09-763-397A-21  
Perfect score: 41  
Sequence: 1 LTPLEELY 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mhc:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phage:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	41	100.0	171	5 Q26002	Q26002 plasmodium
2	41	100.0	171	5 Q26003	Q26003 plasmodium
3	41	100.0	292	5 Q25762	Q25762 plasmodium
4	41	100.0	782	5 Q25730	Q25730 plasmodium
5	41	100.0	782	5 Q26104	Q26104 plasmodium
6	41	100.0	782	5 Q25875	Q25875 plasmodium
7	41	100.0	782	5 Q26007	Q26007 plasmodium
8	41	100.0	782	5 Q9U431	Q9U431 plasmodium
9	41	100.0	782	5 Q9U430	Q9U430 plasmodium
10	41	100.0	782	5 Q9U429	Q9U429 plasmodium
11	41	100.0	782	5 Q9U414	Q9U414 plasmodium
12	39	95.1	446	5 Q9Y0A4	Q9Y0A4 branchiosto
13	38	92.7	444	5 Q21449	Q21449 caenorhabdi
14	37	90.2	448	5 O02441	O02441 geodia cydo
15	36	87.8	149	12 Q9EMU2	Q9EMU2 ansacta moo
16	36	87.8	230	11 Q9CZ18	Q9CZ18 mus musculu
17	36	87.8	448	13 Q93382	Q93382 gallus gall
18	35	85.4	359	5 Q9UAF7	Q9UAF7 branchiosto
19	35	85.4	360	5 Q77158	Q77158 branchiosto

20	35	85.4	443	5 Q9VLB7	Q9VLB7 drosophila
21	35	85.4	448	5 Q24349	Q24349 drosophila
22	34	82.9	276	13 Q9PTH3	Q9PTH3 brachydanio
23	34	82.9	335	2 Q9HY96	Q9HY96 pseudomonas
24	34	82.9	390	2 Q3WY14	Q3WY14 thermotoga
25	34	82.9	412	2 Q9PKD8	Q9PKD8 chlamydia m
26	34	82.9	414	2 O84258	O84258 chlamydia t
27	34	82.9	990	10 Q40699	Q40699 oryza sativ
28	33	80.5	117	2 O07542	O07542 bacillus su
29	33	80.5	139	5 Q9XX29	Q9XX29 caenorhabdi
30	33	80.5	296	2 O53053	O53053 listeria iv
31	33	80.5	338	1 P95904	P95904 sulfolobus
32	33	80.5	392	2 Q99X76	Q99X76 staphylococ
33	33	80.5	393	2 Q9KCN9	Q9KCN9 bacillus ha
34	33	80.5	506	5 Q9V6P1	Q9V6P1 drosophila
35	33	80.5	632	3 Q07845	Q07845 saccharomyc
36	32	78.0	139	10 Q48852	Q48852 arabidopsis
37	32	78.0	139	10 Q9SKY2	Q9SKY2 arabidopsis
38	32	78.0	204	2 O68685	O68685 bacillus ce
39	32	78.0	354	5 Q9TXQ3	Q9TXQ3 caenorhabdi
40	32	78.0	380	3 Q93982	Q93982 zygocacchar
41	32	78.0	387	3 Q9UV50	Q9UV50 debaryomyce
42	32	78.0	393	2 O24821	O24821 bacillus st
43	32	78.0	407	3 Q59854	Q59854 zygocacchar
44	32	78.0	411	2 Q928E9	Q928E9 chlamydia p
45	32	78.0	411	2 Q9JSF2	Q9JSF2 chlamydia p

ALIGNMENTS

RESULT 1  
Q26002  
ID Q26002 PRELIMINARY; PRT; 171 AA.  
AC Q26002;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
DE RHOPTRY PROTEIN (FRAGMENT).  
GN RAP-1.  
OS Plasmodium falciparum.  
OC Eukaryote; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SIERRA LEONE;  
RX MEDLINE=92244303; PubMed=1574089;  
RA Howard R.F.;  
RT "The sequence of the p82 rhoptry protein is highly conserved between two Plasmodium falciparum isolates.";  
RL Mol. Biochem. Parasitol. 51:327-330(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SIERRA LEONE;  
RX MEDLINE=93293317; PubMed=7685740;  
RA Howard R.F., Jensen J.B., Franklin H.L.;  
RT "Reactivity profile of human anti-82-kilodalton rhoptry protein antibodies generated during natural infection with Plasmodium falciparum.";  
RL Infect. Immun. 61:2960-2965(1993).  
DR EMBL; L10322; AAB59202.1; -;  
FT NON\_TER 1  
FT NON\_TER 171  
SQ SEQUENCE 171 AA; 18510 MW; 8BF5709D8D3DB844 CRC64;

Query Match 100.0%; Score 41; DB 5; Length 171;  
Best Local Similarity 100.0%; Pred. No. 1-2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 LTPLEELY 8  
Db 79 LTPLEELY 86

Mon Feb 4 15:23:44 2002

```

RL Mol. Biochem. Parasitol. 77:95-98(1996).
DR EMBL; U41074; AAC47090.1; -.
FT NON_TER 1
SQ SEQUENCE 292 AA; 32345 MW; A35C979C3FEC3019 CRC64;

Query Match
Best Local Similarity 100.0%; Score 41; DB 5; Length 292;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTPLEELY 8
Db 196 LTPLEELY 203

RESULT 4
Q25730 PRELIMINARY; PRT; 782 AA.
AC Q25730;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE RHOPTRY ASSOCIATED PROTEIN-1.
GN RAP-1.
OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FC27;
RA Shi Y., Lal A.A.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20985; AAA63681.1; -.
SQ SEQUENCE 782 AA; 90023 MW; F69E26A2A564C8EA CRC64;

Query Match
Best Local Similarity 100.0%; Score 41; DB 5; Length 782;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTPLEELY 8
Db 202 LTPLEELY 209

RESULT 5
Q26104 PRELIMINARY; PRT; 782 AA.
AC Q26104;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE RHOPTRY ASSOCIATED PROTEIN 1.
GN RAP-1.
OS Plasmodium reichenowi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5854;
RN [1]
RP SEQUENCE FROM N.A.
RA Shi Y., Lal A.A.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20986; AAA63682.1; -.
SQ SEQUENCE 782 AA; 90160 MW; DE6D1BE2FAC308A9 CRC64;

Query Match
Best Local Similarity 100.0%; Score 41; DB 5; Length 782;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTPLEELY 8
Db 202 LTPLEELY 209
```

---

```

RESULT 2
Q26003 PRELIMINARY; PRT; 171 AA.
AC Q26003;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE RHOPTRY PROTEIN (FRAGMENT).
GN RAP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TANZANIA I/CDC;
RX MEDLINE=92244303; PubMed=1574089;
RA Howard R.F.;
RT "The sequence of the p82 rhostry protein is highly conserved between
two Plasmodium falciparum isolates."
RL Mol. Biochem. Parasitol. 51:327-330(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TANZANIA I/CDC;
RX MEDLINE=92293317; PubMed=7685740;
RA Howard R.F., Jensen J.B., Franklin H.L.;
RT "Reactivity profile of human anti-82-kilodalton rhostry protein
antibodies generated during natural infection with Plasmodium
falciparum."
RL Infect. Immun. 61:2960-2965(1993).
DR EMBL; L10323; AAA29742.1; -.
FT NON_TER 1
FT NON_TER 171
SQ SEQUENCE 171 AA; 18531 MW; 192CBAFC68970CB8 CRC64;

Query Match
Best Local Similarity 100.0%; Score 41; DB 5; Length 171;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTPLEELY 8
Db 79 LTPLEELY 86

RESULT 3
Q25762 PRELIMINARY; PRT; 292 AA.
AC Q25762;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE STRAIN INDD RHOPTRY-ASSOCIATED PROTEIN 1 (RAP-1) (FRAGMENT).
GN RAP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INDD (INDIAN ISOLATE D);
RX MEDLINE=92244303; PubMed=1574089;
RA Howard R.F.;
RT "The sequence of the p82 rhostry protein is highly conserved between
two Plasmodium falciparum isolates."
RL Mol. Biochem. Parasitol. 51:327-330(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=INDD (INDIAN ISOLATE D);
RX MEDLINE=96379224; PubMed=8784775;
RA Howard R.F., Peterson C.;
RT "Limited RAP-1 sequence diversity in field isolates of Plasmodium
falciparum."
```



```
RESULT 6
Q25875 ID Q25875 PRELIMINARY; PRT; 782 AA.
AC Q25875;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE PR86 RHOPTRY PRECURSOR PROTEIN.
GN P82.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HONDURAS I CDC;
RX MEDLINE=90348711; PubMed=2200961;
RA Ridley R.G., Takacs B., Lahm H.W., Delves C.J., Goman M., Certa V.,
Matile H., Woollett G.R., Scaife J.G.;
RT "Characterisation and sequence of a protective rhothry antigen from
Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 41:125-134(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HONDURAS I CDC;
RX MEDLINE=92244303; PubMed=1574089;
RA Howard R.F.;
RT "The sequence of the p82 rhothry protein is highly conserved between
two Plasmodium falciparum isolates.";
RL Mol. Biochem. Parasitol. 51:327-330(1992).
DR EMBL; M80807; AAA29717.1; -.
SQ SEQUENCE 782 AA; 90096 MW; DIAD099862528D42 CRC64;

Query Match 100.0%; Score 41; DB 5; Length 782;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LTPLEELY 8
Db 202 LTPLEELY 209

RESULT 7
Q26007 ID Q26007 PRELIMINARY; PRT; 782 AA.
AC Q26007;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE RHOPTRY ASSOCIATED PROTEIN 1 (ROP1) PRECURSOR (FRAGMENT).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=90348711; PubMed=2200961;
RA Ridley R.G., Takacs B., Lahm H.W., Delves C.J., Goman M., Certa V.,
Matile H., Woollett G.R., Scaife J.G.;
RT "Characterisation and sequence of a protective rhothry antigen from
Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 41:125-134(1990).
DR EMBL; M32853; AAA29753.1; -.
KW Signal.
FT SIGNAL. 1 22 POTENTIAL.
FT CHAIN 23 >782 POTENTIAL.
FT NON_TER 782 782
SQ SEQUENCE 782 AA; 90080 MW; 2F2E240D40A4C902 CRC64;

Query Match 100.0%; Score 41; DB 5; Length 782;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LTPLEELY 8
Db 202 LTPLEELY 209

RESULT 8
Q9U430 ID Q9U430 PRELIMINARY; PRT; 782 AA.
AC Q9U430;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE RHOPTRY-ASSOCIATED PROTEIN 1 (FRAGMENT).
GN RAPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FUP;
RA Saul A.;
RT "Efficacy of vaccines containing Rhothry-Associated Proteins RAPI and
RAP2 of Plasmodium falciparum in Saimiri boliviensis monkeys.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF205282; AAF23403.1; -.
FT NON_TER 782 782
SQ SEQUENCE 782 AA; 90066 MW; B344948D5806F7DC CRC64;

Query Match 100.0%; Score 41; DB 5; Length 782;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LTPLEELY 8
Db 202 LTPLEELY 209

RESULT 9
Q9U430 ID Q9U430 PRELIMINARY; PRT; 782 AA.
AC Q9U430;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE RHOPTRY-ASSOCIATED PROTEIN 1 (FRAGMENT).
GN RAPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FUP;
RA Saul A.;
RT "Efficacy of vaccines containing Rhothry-Associated Proteins RAPI and
RAP2 of Plasmodium falciparum in Saimiri boliviensis monkeys.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF205283; AAF23404.1; -.
FT NON_TER 782 782
SQ SEQUENCE 782 AA; 90131 MW; A0FCB64529C34E59 CRC64;

Query Match 100.0%; Score 41; DB 5; Length 782;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LTPLEELY 8
Db 202 LTPLEELY 209

RESULT 10
```

OX NCBI\_TaxID=7739;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RA Sedlacek Z., Shimeld S.M., Muenstermann E., Poustka A.;  
 RT "The amphioxus RAB gdp-dissociation inhibitor (GDI) gene is neural-  
 specific: implications for the evolution of function of chordate RAB  
 GDI genes";  
 RT GDI genes";  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: Y18521; CAB46230.1; -.  
 DR HSSP: P21856; LGND.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR002005; Rab\_GDI\_REP.  
 DR Pfam: PF00996; GDI; 1.  
 DR PRINTS: PR00891; RABGDIREP.  
 DR PROSITE: PS00290; IG\_MHC; UNKNOWN\_1.  
 SQ SEQUENCE 446 AA; 49759 MW; 69256AFFC0188D3A CRC64;

Query Match 95.1%; Score 39; DB 5; Length 446;  
 Best Local Similarity 87.5%; Pred. No. 8.9;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTPLEELY 8  
 Db 202 LTPLEELY 209

Query Match 100.0%; Score 41; DB 5; Length 782;  
 Best Local Similarity 100.0%; Pred. No. 6.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTPLEELY 8  
 Db 202 LTPLEELY 209

RESULT 11  
 Q9U414 PRELIMINARY; PRT; 782 AA.  
 ID Q9U414;  
 AC Q9U414;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
 DE RHOTRY ASSOCIATED PROTEIN-1.  
 GN RAP-1.  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5833;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FCCL/HN;  
 RA Li X.R., Luo S.H., Yu X.B., Shan Z.X., Ma C.L.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF206631; AAF15365.1; -.  
 SQ SEQUENCE 782 AA; 90082 MW; 8E1F4CF2883903FD CRC64;

Query Match 100.0%; Score 41; DB 5; Length 782;  
 Best Local Similarity 100.0%; Pred. No. 6.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTPLEELY 8  
 Db 202 LTPLEELY 209

RESULT 12  
 Q9Y0A4 PRELIMINARY; PRT; 446 AA.  
 ID Q9Y0A4;  
 AC Q9Y0A4;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE RAB GDP-DISSOCIATION INHIBITOR.  
 GN RAB-GDI.  
 OS Branchiostoma floridae (Florida lancelet) (Amphioxus).  
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;  
 OC Branchiostoma.

OX NCBI\_TaxID=7739;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RA Sedlacek Z., Shimeld S.M., Muenstermann E., Poustka A.;  
 RT "The amphioxus RAB gdp-dissociation inhibitor (GDI) gene is neural-  
 specific: implications for the evolution of function of chordate RAB  
 GDI genes";  
 RT GDI genes";  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: Y18521; CAB46230.1; -.  
 DR HSSP: P21856; LGND.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR002005; Rab\_GDI\_REP.  
 DR Pfam: PF00996; GDI; 1.  
 DR PRINTS: PR00891; RABGDIREP.  
 DR PROSITE: PS00290; IG\_MHC; UNKNOWN\_1.  
 SQ SEQUENCE 446 AA; 49759 MW; 69256AFFC0188D3A CRC64;

Query Match 95.1%; Score 39; DB 5; Length 446;  
 Best Local Similarity 87.5%; Pred. No. 8.9;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTPLEELY 8  
 Db 46 LTPLEELY 53

RESULT 13  
 Q21449 PRELIMINARY; PRT; 444 AA.  
 ID Q21449;  
 AC Q21449;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE RABGDI PROTEIN.  
 GN GDI-1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Boguski M.S.;  
 RL Submitted (APR-1993) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RN SEQUENCE FROM N.A.  
 RA McMurray A.;  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U00002; AAA17051.1; -.  
 DR EMBL: Z99281; CAB16511.1; -.  
 DR HSSP: P21856; LGND.  
 DR InterPro: IPR002005; Rab\_GDI\_REP.  
 DR Pfam: PF00996; GDI; 1.  
 DR PRINTS: PR00891; RABGDIREP.  
 SQ SEQUENCE 444 AA; 50025 MW; 0F3CF46B75161DD CRC64;

Query Match 92.7%; Score 38; DB 5; Length 444;  
 Best Local Similarity 87.5%; Pred. No. 14;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LTPLEELY 8  
 Db 46 LTPLEELY 53

RESULT 14  
 O02441 PRELIMINARY; PRT; 448 AA.  
 ID O02441;  
 AC O02441;  
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE GDP-DISSOCIATION INHIBITOR.  
 GN GDI.  
 OS Geodia cydonium (Sponge).  
 OC Eukaryota; Metazoa; Porifera; Demospongiae; Tetractinomorpha;  
 OC Astrophorida; Geodiidae; Geodia.  
 OX NCBI\_TaxID=6047;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Krasko A., Scheffer U., Koziol C., Pancer Z., Batel R., Badria F.A.,  
 RA Mueller W.E.G.; 37:157-168(1997).  
 RL Aquatic Toxicol. 37:157-168(1997).  
 DR EMBL; X94983; CAA64439.1; -.  
 DR HSSP; P21856; ICND.  
 DR InterPro; IPR002005; Rab\_GDI\_REP.  
 DR Pfam; PF00996; GDI; 1  
 DR PRINTS; PR00891; RABGDIREP.  
 DR SEQUENCE 448 AA; 50215 MW; 0F94A980DC68E1A2 CRC64;

Query Match 90.2%; Score 37; DB 5; Length 448;  
 Best Local Similarity 87.5%; Pred. No. 23;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LTPLEELY 8  
 :|:|:|:|:|  
 Db 46 LSPLEELY 53

## RESULT 15

Q9EMU2  
 ID Q9EMU2 PRELIMINARY; PRT; 149 AA.  
 AC Q9EMU2;  
 DT 01-MAR-2001 (TReMBLrel. 16, Created)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
 DE AMV107.  
 GN Ansv107.  
 OS Ansv107.  
 OS Ansv107.  
 OC Ansv107.  
 OC Ansv107.  
 OC Ansv107.  
 OC Ansv107.  
 OX NCBI\_TaxID=28321;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-20396580; PubMed=10936094;  
 RA Bawden A.L., Glassberg K.J., Diggins J., Shaw R., Farmerie W.,  
 RA Moyer R.W.;  
 RA "Complete Genomic Sequence of the Ansv107 Entomopoxvirus:  
 RT Analysis and Comparison with Other Poxviruses.";  
 RL Virology 274:120-139(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Bawden A.L., Glassberg K.J., Diggins J., Shaw R., Farmerie W.,  
 RA Moyer R.W.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF250284; AAG02813.1; -.  
 SQ SEQUENCE 149 AA; 17470 MW; 877185CEA034F840 CRC64;

Query Match 87.8%; Score 36; DB 12; Length 149;  
 Best Local Similarity 75.0%; Pred. No. 11;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LTPLEELY 8  
 :|:|:|:|:|  
 Db 85 ITPVEELY 92

Search completed: January 29, 2002, 11:12:16  
 Job time: 772 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 10:21:56 ; Search time 66.28 seconds  
(without alignments)  
18.999 Million cell updates/sec

Title: US-09-763-397A-22

Perfect score: 79

Sequence: 1 SSPSTKSPSNVKSAS 17

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_ll01.\*  
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	75	94.9	16	21	AAV70298		plasmidium falcipa
2	68.5	86.7	350	21	AAV70278		Recombinant vaccin
3	68.5	86.7	782	11	AAV06991		Polypeptide antige
4	49	62.0	530	13	AAV28888		R1 paraneoplastic
5	48	60.8	140	22	AAV98924		Arabidopsis thalia
6	48	60.8	328	21	AAV54617		Arabidopsis Abi4 p
7	48	60.8	328	22	AAV98923		Arabidopsis thalia
8	48	60.8	328	22	AAE01347		Arabidopsis thalia
9	46	58.2	139	21	AAV23324		Arabidopsis thalia
10	46	58.2	141	21	AAV23323		Arabidopsis thalia
11	46	58.2	147	21	AAV23322		Arabidopsis thalia

12	46	58.2	243	22	AAU01602	Human secreted pro
13	46	58.2	273	21	AAV70705	Maize replication
14	46	58.2	273	21	AAV70706	Maize replication
15	46	58.2	273	21	AAV70707	Maize replication
16	46	58.2	273	21	AAV70709	Maize replication
17	46	58.2	273	21	AAV70710	Maize replication
18	46	58.2	303	22	AAV70710	Human polypeptide
19	46	58.2	310	22	AAV70710	Human polypeptide
20	46	58.2	310	22	AAV70710	Human polypeptide
21	46	58.2	310	22	AAV70710	Human polypeptide
22	46	58.2	310	22	AAV70710	Human polypeptide
23	46	58.2	310	22	AAV70710	Human polypeptide
24	46	58.2	310	22	AAV70710	Human polypeptide
25	46	58.2	311	22	AAV70710	Human polypeptide
26	46	58.2	318	21	AAV70708	Maize replication
27	46	58.2	473	21	AAV70708	Maize replication
28	45	57.0	30	17	AAV70708	Lysine/Serine cont
29	45	57.0	75	18	AAV70708	Peptide carrier #2
30	45	57.0	205	22	AAV70708	Human colon cancer
31	45	57.0	583	21	AAV70708	Human BMAL1a trans
32	45	57.0	626	21	AAV70708	Human BMAL1b trans
33	45	57.0	626	21	AAV70708	Human JAP3 transcr
34	45	57.0	2559	20	AAV70708	Amino acid sequenc
35	44	55.7	176	22	AAV70708	Arabidopsis thalia
36	44	55.7	197	21	AAV70708	Human polypeptide
37	44	55.7	244	22	AAV70708	Human polypeptide
38	44	55.7	285	22	AAV70708	Human protein sequ
39	44	55.7	817	22	AAV70708	Human polypeptide
40	44	55.7	819	22	AAV70708	Human polypeptide
41	44	55.7	819	22	AAV70708	Human polypeptide
42	44	55.7	863	22	AAV70708	Human polypeptide
43	44	55.7	2432	21	AAV70708	Human homologue of
44	43	54.4	30	17	AAV70708	Lysine/Serine cont
45	43	54.4	33	22	AAV70708	Peptide #6352 enco

#### ALIGNMENTS

RESULT 1  
AAV70298  
ID AAV70298 standard; peptide: 16 AA.  
AC AAV70298;  
DT 06-JUN-2000 (first entry)  
XX Plasmodium falciparum RAP-1 antigenic epitope, P604.  
DE Recombinant protein; CDC/NIAID/VAC-1; multivalent; malaria; vaccine;  
XX T-cell epitope; tetanus toxoid; antigenic epitope; treatment;  
XX Circumsporozoite protein; CSP; sporozoite surface protein-2; SSP-2;  
KW Liver stage antigen-1; LSA-1; merozoite surface protein-1; MSP-1;  
KW Apical membrane antigen-1; AMA-1; erythrocyte binding antigen-175;  
KW EBA-175; rhoptry associated protein-1; RAP-1; gamete specific antigen;  
KW Pf27; antiparasitic; prevention; anti-CDC/NIAID/VAC-1 antibody.  
XX Plasmodium falciparum.  
OS Plasmodium falciparum.  
XX WO200011179-A1.  
XX 02-MAR-2000.  
XX 19-AUG-1999; 99WO-US18869.  
XX 21-AUG-1998; 98US-0097703.  
XX (NAIM-) NAT INST IMMUNOLOGY.  
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.  
XX Lal AA, Shi YP, Hasnain SE;  
XX WPI; 2000-237654/20.

XX  
PT Novel recombinant protein as vaccine for treating malarial infection  
PT comprises antigenic peptides obtained from different stages of  
PT Plasmodium falciparum life cycle -  
XX  
PS Claim 2; Page 17; 52pp; English.  
XX  
XX The present sequence is the antigenic epitope p604, derived from  
XX rhoptry associated protein-1 (RAP-1) of the asexual blood stage of  
CC Plasmodium falciparum. It is used in the construction of recombinant  
CC protein CDC/NiMALVAC-1, which is a multivalent, multistage malarial  
CC vaccine. The recombinant protein comprises, melittin signal peptide,  
CC (His)6 tag, T-cell epitope from tetanus toxoid and 21 antigenic epitopes  
CC from circumsporozoite protein (CSP), sporozoite surface protein-2  
CC (SSP-2), liver stage antigen-1 (LSA-1), merozoite surface protein-1  
CC (MSP-1), MSP-2, apical membrane antigen-1 (AMA-1), erythrocyte binding  
CC antigen-175 (EBA-175), rhoptry associated protein-1 (RAP-1) and gamete  
CC specific antigen, Pf927. These epitopes were obtained at different stages  
CC of the life cycle of P. falciparum. CDC/NiMALVAC-1 vaccine has  
CC antiparasitic activity and can be used for treatment and prevention of  
CC malarial infections. Anti-CDC/NiMALVAC-1 antibodies can be used for  
XX detecting P. falciparum in biological samples.  
XX Sequence 16 AA;  
SQ

Query Match 94.9%; Score 75; DB 21; Length 16;  
Best Local Similarity 100.0%; Pred. No. 4.5e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSPSTKSSPSNVKSA 16  
Db 1 sspstsksspsnvksa 16  
|||||

RESULT 2  
AA70278  
ID AAY70278 standard; Protein; 350 AA.  
XX  
AC AAY70278;  
XX  
XX 06-JUN-2000 (first entry)  
XX  
XX Recombinant vaccine CDC/NiMALVAC-1.  
XX  
XX Recombinant protein; CDC/NiMALVAC-1; multivalent; malaria; vaccine;  
KW T-cell epitope; tetanus toxoid; antigenic epitope; treatment;  
KW circumsporozoite protein; CSP; sporozoite surface protein-2; SSP-2;  
KW liver stage antigen-1; LSA-1; merozoite surface protein-1; MSP-1; MSP-2;  
KW apical membrane antigen-1; AMA-1; erythrocyte binding antigen-175;  
KW EBA-175; rhoptry associated protein-1; RAP-1; gamete specific antigen;  
KW Pf927; antiparasitic; prevention; anti-CDC/NiMALVAC-1 antibody;  
KW honey bee.  
XX  
XX Chimeric - Apis sp.  
OS Chimeric - Clostridium tetani.  
OS Chimeric - Plasmodium falciparum.  
XX  
XX Key Location/Qualifiers  
FT Peptide 1..22  
FT /label= Melittin\_signal\_peptide  
FT /note= "Derived from Honey bee"  
FT Protein 23..350  
FT /label= Mature\_CDC/NiMALVAC-1  
FT /note= "Recombinant multivalent malarial vaccine"  
XX  
XX WO200011179-A1.  
XX  
XX 02-MAR-2000.  
XX  
XX 19-AUG-1999; 99WO-US18869.  
XX  
XX 21-AUG-1998; 98US-0097703.  
XX  
XX

XX  
PA (NATM-) NAT INST IMMUNOLOGY.  
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.  
XX  
XX Lal AA, Shi YP, Hasnain SE;  
XX  
XX WPI: 2000-237654/20.  
XX N-PSDB; AA051336.  
XX  
XX Novel recombinant protein as vaccine for treating malarial infection  
PT comprises antigenic peptides obtained from different stages of  
PT Plasmodium falciparum life cycle -  
XX  
XX Claim 3; Page 43-44; 52pp; English.  
XX  
XX The present sequence is that of recombinant protein CDC/NiMALVAC-1,  
CC which is a multivalent, multistage malarial vaccine. The recombinant  
CC protein comprises, melittin signal peptide, (His)6 tag, T-cell epitope  
CC from tetanus toxoid and 21 antigenic epitopes from circumsporozoite  
CC protein (CSP), sporozoite surface protein-2 (SSP-2), liver stage  
CC antigen-1 (LSA-1), merozoite surface protein-1 (MSP-1), MSP-2, apical  
CC membrane antigen-1 (AMA-1), erythrocyte binding antigen-175 (EBA-175),  
CC rhoptry associated protein-1 (RAP-1) and gamete specific antigen, Pf927.  
CC These epitopes were obtained at different stages of the life cycle of  
CC Plasmodium falciparum. CDC/NiMALVAC-1 vaccine has antiparasitic  
CC activity and can be used for treatment and prevention of malarial  
CC infections. Anti-CDC/NiMALVAC-1 antibodies can be used for detecting  
XX P. falciparum in biological samples.  
XX Sequence 350 AA;  
SQ

Query Match 86.7%; Score 68.5; DB 21; Length 350;  
Best Local Similarity 94.4%; Pred. No. 0.012;  
Matches 17; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 SSPSTKSSPSNVKSA 17  
Db 299 sspstsksspsnvksas 316  
|||||

RESULT 3  
AAR06991  
ID AAR06991 standard; protein; 782 AA.  
XX  
XX AAR06991;  
XX  
XX 16-JAN-1991 (first entry)  
XX  
XX Polypeptide antigenic to rhoptry organelles of Plasmodium.  
DE  
XX  
XX Malaria; vaccine; rhoptry organelles.  
KW  
XX  
XX Plasmodium falciparum.  
OS  
XX  
XX EF388738-A.  
FN  
XX  
XX 26-SEP-1990.  
PD  
XX  
XX 09-MAR-1990; 90EP-0104561.  
PF  
XX  
XX 22-AUG-1989; 89GB-0019064.  
PR  
XX  
XX 14-MAR-1989; 89GB-0005857.  
PR  
XX  
XX (HOFF) HOFFMANN-LA ROCHE AG.  
PA  
XX  
XX Ridley RG, Scaife JG;  
PI  
XX  
XX WPI: 1990-291721/39.  
DR  
XX  
XX N-PSDB; AAQ06000.  
DR  
XX  
XX Antigenic polypeptide and DNA encoding it - having a determinant  
PT cross reactive with those on the rhoptry organelles of the  
PT

PT merozoite form of the malaria parasite P falciparum.

PS Claim 1; Fig 2a-b; 29pp; English.

CC Gene product may be isolated from a transformed E.coli (Y1088)  
 CC expression system using plasmid pMC9, and may then be used for  
 CC immunisation against malaria. The product may also be used  
 CC diagnostically to detect Abs directed against the parasite.

XX Sequence 782 AA;

Query Match 86.7%; Score 68.5; DB 11; Length 782;  
 Best Local Similarity 94.4%; Pred. No. 0.03;  
 Matches 17; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 SSPSSTK-SSPSNVKSAS 17  
 Db 145 sspsstkspsnvksas 162  
 ||||| ||||| |||||

RESULT 4  
 AAR2888  
 ID AAR28888 standard; Protein; 530 AA.

XX AC AAR28888;

XX DT 25-MAR-1993 (first entry)

XX DE R1 paraneoplastic antigenic polypeptide.

XX KW R18; paraneoplastic opsoclonus; PO; cerebellar; expression vector;  
 KW library; R1; paraneoplastic antigenic polypeptide; phagemid;  
 KW lambda-R18; helper phage; R408; anti-R1; antiserum; fusion protein;  
 KW antibody; dideoxy method; open reading frame; PCR; probe.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT Region 72..107

FT /note= "HOM #1"

FT Region 197..232

FT /note= "HOM #2"

FT Region 447..482

FT /note= "HOM #3"

XX PN W09219635-A.

XX PD 12-NOV-1992.

XX PF 21-APR-1992; 92WO-US03249.

XX PR 25-APR-1991; 91US-0691559.

XX PA (SLOK ) SLOAN KETTERING INST CANCER.

XX PI Darnell RB, Furneaux HW, Posner JB;

XX DR WPI; 1992-398791/48.

XX DR N-PSDB; AAQ30802.

PT R1 para-neoplastic antigenic polypeptide and antibody specific

PT for para-neoplastic antigen - used in the diagnosis and treatment

PT of para-neoplastic opsoclonus

XX Claim 5; Fig 7; 50pp; English.

XX This sequence is the R1 paraneoplastic antigenic polypeptide. The R1  
 CC gene is associated with paraneoplastic opsoclonus (PO). The DNA  
 CC sequence encoding this polypeptide was isolated via an R1 clone, R18,  
 CC which was isolated from a cerebellar cDNA expression vector library  
 CC derived from antiserum from a single patient with PO. Phagemid was  
 CC excised from lambda-R18 using the helper phage R408, the resulting

CC plasmid being designated pri8 (see also AAQ30801). pri8 was cloned and  
 CC propagated in E. coli. pri8 encoded a protein of mol. wt. 40 kD which  
 CC is recognised by anti-R1 antiserum. This R18 fusion protein was used  
 CC as an antigen to affinity purify antibody from the anti-R1 antiserum.  
 CC The DNA sequence of pri8 was determined by the dideoxy method and an  
 CC open reading frame encoding approx. 340 amino acids was identified.  
 CC Regions of this sequence could be used to make PCR probes such that  
 CC additional clones representing the full length R1 gene could be  
 CC isolated.

XX Sequence 530 AA;

Query Match 62.0%; Score 49; DB 13; Length 530;  
 Best Local Similarity 58.8%; Pred. No. 16;  
 Matches 10; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 SSPSSTKSSPSNVKSAS 17  
 Db 176 sspttksspsqpmmts 192  
 ||||| ||||| : |

RESULT 5

AAB98924  
 ID AAB98924 standard; peptide; 140 AA.

XX AC AAB98924;

XX DT 09-OCT-2001 (first entry)

XX DE Arabidopsis thaliana abi4 allele-encoded protein.

XX KW AB14 gene; transactivator; APETALA2 domain; AP2 domain; chromosome 2;  
 KW C-repeat/DRE regulatory element; abscisic acid insensitivity;  
 KW salt tolerance; seed germination; saline soil; abi4 allele.

XX OS Arabidopsis thaliana.

XX FH Key Location/Qualifiers

FT Domain 55..110

FT /label= APETALA2\_domain  
 FT /note= "Binds to C-repeat/DRE regulatory elements;  
 acts as a transactivator."

XX PN W0200149850-A1.

XX PD 12-JUL-2001.

XX PF 03-JAN-2001; 2001WO-ES00003.

XX PR 03-JAN-2000; 2000ES-0000004.

XX PA (UYEL-) UNIV ELCHE MIGUEL HERNANDEZ.

XX PI Quesada Perez V, Ponce Molet RM, Micol Molina JL;

XX DR WPI; 2001-441881/47.

XX New null allele of the Arabidopsis thaliana AB14 gene, useful for  
 PT imparting insensitivity to abscisic acid and ability to germinate and  
 PT grow on saline soils -

XX PS Disclosure; Fig 5; 26pp; Spanish.

XX The invention relates to a null, or extremely hypomorphic, allele of  
 CC the AB14 gene of Arabidopsis thaliana that confers insensitivity to  
 CC abscisic acid and tolerance to salinity during germination. The allele,  
 CC designated san5 or abi4-2, has a deletion of one of the three adenines at  
 CC positions 277-279 relative to the wild-type allele (designated AB14),  
 CC which causes a change in the reading frame leading to a shorter, variant  
 CC C-terminus from amino acid residue 44 onwards. The abi4-2 protein  
 CC (AAB98925) does not contain the APETALA2 (AP2) domain located between  
 CC residues 55-110 in the wild-type AB14 protein (AAB98923) which mediates

Mon Feb 4 15:23:44 2002

transactivational activity at C-repeat/DRE regulatory elements. Seeds of Arabidopsis thaliana homozygous for the abi4-2 (san5) allele are tolerant of high concentrations of sodium chloride (150 mM), mannitol (400 mM), potassium chloride (150 mM) and sodium sulphate (75 mM). The abi4-2 (san5) allele is used to produce Arabidopsis plants that are able to germinate and grow in highly saline soils in which wild-type plants cannot survive. Orthologous alleles from other crop plants can be used to impart similar properties. The present sequence represents the abi4 protein encoded by the mutant allele abi4. This protein contains the APETALA2 domain, but is C-terminally truncated with respect to the wild-type ABI4 protein.

XX Sequence 140 AA;

Query Match 60.8%; Score 48; DB 22; Length 140;  
Best Local Similarity 64.7%; Pred. No. 5.3;  
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 SSPSTKSSPSNVKSAS 17  
Db 113 sspssvssssvsas 129  
||||| ||| :||

RESULT 6  
AAV54617  
ID AAV54617 standard; Protein; 328 AA.

AC AAV54617;

DT 02-FEB-2000 (first entry)

DE Arabidopsis Abi4 protien sequence.

KW Abi4 gene; seed development regulation; seed nutrient reserve production;  
KW desiccation protectant; vegetative stress tolerance.

OS Arabidopsis thaliana.

PN W09955840-A1.

PD 04-NOV-1999.

PF 27-APR-1999; 99WO-US08954.

PR 27-APR-1998; 98US-0083334.

XX (REGC ) UNIV CALIFORNIA.  
XX (GEO ) GEN HOSPITAL CORP.

DR WPI: 2000-013434/01.

DR N-PSDB; AA237307, AA237308.

XX Arabidopsis thaliana ABI4 gene, mutants and variants, useful for  
XX creating transgenic plants with e.g. modified regulation of seed  
XX viability and vegetative stress tolerance

PS Claim 6; Fig 4; 68pp; English.

XX This sequence represents the Arabidopsis thaliana Abi4 protein of the  
XX invention. Abi4 is involved in regulating seed development, production of  
XX seed nutrient reserves and desiccation protectants. The nucleic acid  
XX can be used to create transgenic plants with modified regulation of seed  
XX viability, production of nutrient reserves and desiccation protectants in  
XX seeds and vegetative stress tolerance, e.g. to drought or salinity, and  
XX response to abscisic acid.

XX Sequence 328 AA;

Query Match 60.8%; Score 48; DB 21; Length 328;  
Best Local Similarity 64.7%; Pred. No. 13;  
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 SSPSTKSSPSNVKSAS 17  
Db 113 sspssvssssvsas 129  
||||| ||| :||

RESULT 7

ID AAB98923 standard; peptide; 328 AA.

XX AAB98923;

XX 09-OCT-2001 (first entry)

XX Arabidopsis thaliana ABI4 wild-type protein.

XX ABI4 gene; transactivator; APETALA2 domain; AP2 domain; chromosome 2;  
KW C-repeat/DRE regulatory element; abscisic acid insensitivity;  
KW salt tolerance; seed germination; saline soil; wild-type allele.

XX Arabidopsis thaliana.

XX Key Location/Qualifiers

XX Domain 55..110

XX /label= APETALA2 domain

XX /note= "Binds to C-repeat/DRE regulatory elements;  
XX acts as a transactivator"

XX W0200149850-A1.

XX 12-JUL-2001.

XX 03-JAN-2001; 2001WO-ES000003.

XX 03-JAN-2000; 2000ES-00000004.

XX (UYEL-) UNIV ELCHE MIGUEL HERNANDEZ.

XX Quesada Perez V, Ponce Molet RM, Micol Molina JL;

XX WPI: 2001-441881/47.

XX New null allele of the Arabidopsis thaliana ABI4 gene, useful for  
XX imparting insensitivity to abscisic acid and ability to germinate and  
XX grow on saline soils

XX Disclosure; Fig 5; 26pp; Spanish.

XX The invention relates to a null, or extremely hypomorphic, allele of  
XX the ABI4 gene of Arabidopsis thaliana that confers insensitivity to  
XX abscisic acid and tolerance to salinity during germination. The allele,  
XX designated san5 or abi4-2, has a deletion of one of the three adenines at  
XX positions 277-279 relative to the wild-type allele (designated ABI4),  
XX which causes a change in the reading frame leading to a shorter, variant  
XX C-terminus from amino acid residue 44 onwards. The abi4-2 protein  
XX (AAB98923) does not contain the APETALA2 (AP2) domain located between  
XX residues 55-110 in the wild-type ABI4 protein (AAB98923) which mediates  
XX transactivational activity at C-repeat/DRE regulatory elements. Seeds of  
XX Arabidopsis thaliana homozygous for the abi4-2 (san5) allele are tolerant  
XX of high concentrations of sodium chloride (150 mM), mannitol (400 mM),  
XX potassium chloride (150 mM) and sodium sulphate (75 mM). The abi4-2  
XX (san5) allele is used to produce Arabidopsis plants that are able to  
XX germinate and grow in highly saline soils in which wild-type plants  
XX cannot survive. Orthologous alleles from other crop plants can be used  
XX to impart similar properties. The present sequence represents the  
XX wild-type ABI4 protein.

XX Sequence 328 AA;

Query Match 60.8%; Score 48; DB 22; Length 328;  
Best Local Similarity 64.7%; Pred. No. 13;  
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;



```
Oy 1 SSPSSTKSSPSNVKSAS 17
    ||||| ||| :||
Db 113 sspssvsssssvsaas 129

RESULT 8
AAE01347
ID AAE01347 standard; Protein; 328 AA.
XX AC AAE01347;
XX AC AAG23324;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 26590.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EPI033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 28-APR-1999; 99US-0130891.
XX PR 30-APR-1999; 99US-0131449.
XX PR 30-APR-1999; 99US-0132048.
XX PR 30-APR-1999; 99US-0132407.
XX PR 04-MAY-1999; 99US-0132484.
XX PR 05-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 06-MAY-1999; 99US-0132487.
XX PR 07-MAY-1999; 99US-0132863.
XX PR 11-MAY-1999; 99US-0134256.
XX PR 14-MAY-1999; 99US-0134218.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134221.
XX PR 14-MAY-1999; 99US-0134370.
XX PR 18-MAY-1999; 99US-0134768.
XX PR 19-MAY-1999; 99US-0134941.
XX PR 20-MAY-1999; 99US-0135124.
XX PR 21-MAY-1999; 99US-0135353.
XX PR 24-MAY-1999; 99US-0135629.
XX PR 25-MAY-1999; 99US-0136021.
XX PR 27-MAY-1999; 99US-0136392.
XX PR 28-MAY-1999; 99US-0136782.
XX PR 01-JUN-1999; 99US-0137222.
XX PR 03-JUN-1999; 99US-0137528.
XX PR 04-JUN-1999; 99US-0137502.
XX PR 07-JUN-1999; 99US-0137724.
XX PR 08-JUN-1999; 99US-0138094.
XX PR 10-JUN-1999; 99US-0138540.
XX PR 10-JUN-1999; 99US-0138847.
XX PR 14-JUN-1999; 99US-0139119.
XX PR 16-JUN-1999; 99US-0139452.
XX PR 16-JUN-1999; 99US-0139453.
XX PR 17-JUN-1999; 99US-0139492.
XX PR 18-JUN-1999; 99US-0139454.
XX PR 18-JUN-1999; 99US-0139455.
XX PR 18-JUN-1999; 99US-0139456.
XX PR 18-JUN-1999; 99US-0139457.
XX PR 18-JUN-1999; 99US-0139458.
XX PR 18-JUN-1999; 99US-0139459.

Oy 1 SSPSSTKSSPSNVKSAS 17
    ||||| ||| :||
Db 113 sspssvsssssvsaas 129

Query Match 60.8%; Score 48; DB 22; Length 328;
Best Local Similarity 64.7%; Pred. No. 13;
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

RESULT 9
AAE01347
ID AAE01347 standard; Protein; 328 AA.
XX AC AAE01347;
XX AC AAG23324;
XX DT 31-JUL-2001 (first entry)
XX DE Arabidopsis thaliana ABA (abscisic acid)-insensitive 4, ABI4 protein.
XX KW Abscisic acid; ABA; ABA-insensitive; ABI4; signal transduction;
XX KW kernel embryo promoter; cereal plant; maize; rice; cotton; citrus tree;
XX KW tomato; pine; soybean; peanut; olive.
XX OS Arabidopsis thaliana.
XX PN WO200136596-A2.
XX PD 25-MAY-2001.
XX PF 17-NOV-2000; 2000WO-US31739.
XX PR 17-NOV-1999; 99US-0156080.
XX FA (PION-) PIONEER HI-BRED INT INC.
XX PI Helentjaris T;
XX WPI: 2001-329218/34.
XX N-PSDB: AAD05663.
XX DNA constructs containing an abscisic acid-associated sequence are used
XX PT to modulate a response in a plant to abscisic acid so that the plants
XX PT are protected against any harmful or detrimental effects of stress and
XX PT environmental conditions -
XX PS Claim 3; Page 57-58; 60pp; English.
XX CC The invention relates to a method for modulating a response in a target
XX CC plant to abscisic acid (ABA) by introducing a DNA construct comprising
XX CC an ABA-associated sequence operably linked to an early kernel/embryo
XX CC promoter into the plant. The method also involves modulating ABA
XX CC perception and signal transduction in developing seeds. The method is
XX CC used to produce plants, particularly cereal plants such as maize, which
XX CC have a modified response to ABA so that an increase in yield can be
XX CC produced. The plants are protected against any harmful/detrimental
XX CC effects of stress and adverse environmental conditions. The ABA function
XX CC is disrupted in a tissue and developmental preferred manner so that
XX CC female reproductive tissue is insulated from the stress or adverse
XX CC environmental conditions. Other plants which can be transformed include
XX CC rice, cotton, citrus trees, tomatoes, pines, soybean, peanut and olive.
XX CC The present sequence is Arabidopsis thaliana ABA-insensitive 4 (ABI4)
XX CC protein.
XX SQ Sequence. 328 AA;
```



DE Arabidopsis thaliana protein fragment SEQ ID NO: 26589.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-01343768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 10-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.

```

PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 08-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 08-OCT-1999; 99US-0158369.
PR 12-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 58.2%; Score 46; DB 21; Length 141;
Best Local Similarity 56.2%; Pred. NO. 11;
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 SPSTKSPSNVKAS 17
    ||||| |::| |
Db 71 spspsdpts1kcaa 86

RESULT 11
AAG23322
ID AAG23322 standard; Protein; 147 AA.
XX AC
XX AC AAG23322;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 26588.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX OS
```

```

XX EP1033405-A2.
XX PN
XX XX
PD 06-SEP-2000.
XX PF
XX 25-FEB-2000; 2000EP-0301439.
XX XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
```

PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 03-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.

PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 58.2%; Score 46; DB 21; Length 147;

Best Local Similarity 56.2%; Pred. No. 11;

Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 SPSTKSPSNVKSAS 17

||||| |::| |:

Db 77 spstpsdptsikcaa 92

RESULT 12

AAU01602

ID AAU01602 standard; Protein; 243 AA.

XX

AC AAU01602;

XX

DT 18-JUL-2001 (first entry)

XX

DE Human secreted protein encoded by gene #16.

XX

KW Human secreted protein; autoimmune disorder; hyperproliferative disorder;

KW cardiovascular disorder; cerebrovascular disorder; angiogenesis;

KW nervous system disorder; bacterial infection; viral infection;

KW fungal infection; ocular disorder; wound healing; tissue regeneration;

KW epithelial cell proliferation; skin ageing; chemotaxis; IgG Fc region.

XX Homo sapiens.

OS

XX WO200123547-A1.

PN

XX 05-APR-2001.

PD

XX

Mon Feb 4 15:23:44 2002

PF 26-SEP-2000; 2000WO-US26337.  
 XX  
 PR 27-SEP-1999; 99US-0155806.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Komatsoulis GA, Ruben SM, Rosen CA;  
 XX  
 DR WPI; 2001-266151/27.  
 XX N-PSDB; AAS02655.  
 XX  
 PT Nucleic acids encoding 26 human secreted polypeptides, useful for  
 PT preventing, diagnosing and/or treating e.g. Gaucher's disease,  
 PT Alzheimer's disease, Schmiter syndrome, Creutzfeldt-Jacob disease,  
 PT diabetes mellitus and multiple sclerosis -  
 XX  
 PS Disclosure; Page 399; 412pp; English.  
 XX  
 CC The sequence represents a human secreted protein encoded by a nucleic  
 CC acid of the invention. Secreted proteins and their related nucleic acids  
 CC can be used in the diagnosis of or susceptibility to a pathological  
 CC condition by determining the presence or absence of a mutation in a  
 CC nucleic acid or the presence or amount of expression of a secreted  
 CC protein. The sequences are used to prevent, treat or ameliorate a medical  
 CC condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs,  
 CC chickens or sheep. The antibodies to the polypeptides can also be used in  
 CC alleviating symptoms associated with disorders and in diagnostic  
 CC immunoassays e.g. radioimmunoassays or enzyme linkedCC immunosorbent  
 CC assays (ELISA). The disorders include autoimmune diseases e.g. rheumatoid  
 CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or  
 CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular  
 CC disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders  
 CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and  
 CC fungi and ocular disorders e.g. corneal infection. The peptides can also  
 CC be used to aid wound healing and epithelial cell proliferation, to help  
 CC prevent skin ageing due to sunburn, to maintain organs before  
 CC transplantation, to regenerate tissues, in chemotaxis and as a food  
 CC additive or preservative to alter storage capabilities.  
 XX  
 SQ Sequence 243 AA;

Query Match 58.2%; Score 46; DB 22; Length 243;  
 Best Local Similarity 64.7%; Pred. No. 19;  
 Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 SSPSTKSSPSNVKAS 17  
 ||||| :|| | |||  
 Db 156 sspsslstsppevfis 172

RESULT 13  
 AAY70705  
 ID AAY70705 standard; Protein; 273 AA.  
 XX  
 AC AAY70705;  
 XX  
 DT 18-JUL-2000 (first entry)  
 XX  
 DE Maize replication protein A middle subunit homologue-1 (ZMRPAMSH1).  
 XX  
 KW Maize; Replication protein A; RPA; middle subunit; ZMRPAMSH; virucide;  
 KW fungicide; insecticide; chromosome 5; DNA-binding protein; DNA repair;  
 KW DNA metabolism; DNA replication; cell cycle; homologous recombination;  
 KW pathogen resistance; fungi; virus; nematode; insect; gene therapy;  
 KW genetic manipulation.  
 XX  
 OS Zea mays.  
 XX  
 PN WO200015816-A2.  
 XX  
 PD 23-MAR-2000.  
 XX  
 PF 15-SEP-1999; 99WO-US21277.  
 XX  
 PR 17-SEP-1998; 98US-0100690.  
 PR 11-MAR-1999; 99US-0123896.  
 XX  
 PA (PION-) PIONEER HI-BRED INT INC.  
 XX  
 PI Mahajan P;

PF 15-SEP-1999; 99WO-US21277.  
 XX  
 PR 17-SEP-1998; 98US-0100690.  
 PR 11-MAR-1999; 99US-0123896.  
 XX  
 PA (PION-) PIONEER HI-BRED INT INC.  
 XX  
 PI Mahajan P;  
 XX  
 DR WPI; 2000-271452/23.  
 DR N-PSDB; AAZ52287.  
 XX  
 PT New maize replication protein A useful for genetic transformation, gene  
 PT targeting in plants and modulating DNA metabolism -  
 XX  
 PS Claim 1c; Page 88; 101pp; English.  
 XX  
 CC The present sequence is the maize replication protein A (RPA) middle  
 CC subunit homologue-1 (ZMRPAMSH1). RPA is a single-stranded DNA-binding  
 CC protein required for multiple processes in DNA metabolism, like DNA  
 CC replication, repair mechanism (e.g. nucleotide excision and double  
 CC stranded (ds) DNA break repair) and recombination. The gene for the  
 CC ZMRPAMSH homologue is mapped to chromosome 5. This sequence has  
 CC fungicide, virucide and insecticidal activity. DNA encoding RPA is  
 CC used for modulating DNA metabolism, influencing cell cycle, enhancing  
 CC homologous recombination and increasing pathogen resistance in plants.  
 CC Pests that can be controlled include fungal pathogens, viruses,  
 CC nematodes and insects. Antisense sequences can be used to block RPA  
 CC expression and promote non-specific recombination events. RPA protein  
 CC can be used to improve genetic manipulation and also in gene therapy.  
 XX  
 SQ Sequence 273 AA;

Query Match 58.2%; Score 46; DB 21; Length 273;  
 Best Local Similarity 58.8%; Pred. No. 22;  
 Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 SSPSTKSSPSNVKAS 17  
 |::| |::| |::| |::|  
 Db 190 stptsikssppvts 206

RESULT 14  
 AAY70706  
 ID AAY70706 standard; Protein; 273 AA.  
 XX  
 AC AAY70706;  
 XX  
 DT 18-JUL-2000 (first entry)  
 XX  
 DE Maize replication protein A middle subunit homologue-2 and 3.  
 XX  
 KW Maize; Replication protein A; RPA; middle subunit; ZMRPAMSH; virucide;  
 KW fungicide; insecticide; chromosome 5; DNA-binding protein; DNA repair;  
 KW DNA metabolism; DNA replication; cell cycle; homologous recombination;  
 KW pathogen resistance; fungi; virus; nematode; insect; gene therapy;  
 KW genetic manipulation.  
 XX  
 OS Zea mays.  
 XX  
 PN WO200015816-A2.  
 XX  
 PD 23-MAR-2000.  
 XX  
 PF 15-SEP-1999; 99WO-US21277.  
 XX  
 PR 17-SEP-1998; 98US-0100690.  
 PR 11-MAR-1999; 99US-0123896.  
 XX  
 PA (PION-) PIONEER HI-BRED INT INC.  
 XX  
 PI Mahajan P;

XX WPI; 2000-271452/23.  
 DR N-PSDB; AAZ52288.  
 XX  
 PT New maize replication protein A useful for genetic transformation, gene  
 PT targeting in plants and modulating DNA metabolism -  
 XX  
 PS Claim 1c; Page 90-91; 101pp; English.

XX The present sequence is the maize replication protein A middle subunit  
 CC homologue-2 and 3 (ZMRPAMSH2 and ZMRPAMSH3). RPA is a single-stranded  
 CC DNA-binding protein required for multiple processes in DNA metabolism,  
 CC like DNA replication, repair mechanism (e.g. nucleotide excision and  
 CC double stranded (ds) DNA break repair) and recombination. The gene for  
 CC the ZMRPAMS homologue is mapped to chromosome 5. This sequence has  
 CC fungicide, virucide and insecticidal activity. DNA encoding RPA is used  
 CC for modulating DNA metabolism, influencing cell cycle, enhancing  
 CC homologous recombination and increasing pathogen resistance in plants.  
 CC Pests that can be controlled include fungal pathogens, viruses,  
 CC nematodes and insects. Antisense sequences can be used to block RPA  
 CC expression and promote non-specific recombination events. RPA protein  
 CC can be used to improve genetic manipulation and also in gene therapy.

XX Sequence 273 AA;

Query Match 58.2%; Score 46; DB 21; Length 273;  
 Best Local Similarity 58.8%; Pred. No. 22;  
 Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 SSPSSTKSSPSNVKSAS 17  
 I:|:| ||||: | | |  
 Db 190 stptslksspapvtsgs 206

# RESULT 15

AAAY70707  
 ID AAY70707 standard; Protein: 273 AA.

AC AAY70707;

DT 18-JUL-2000 (first entry)

XX Maize replication protein A middle subunit homologue-4 (ZMRPAMSH4).

XX Maize: Replication protein A; RPA; middle subunit; ZMRPAMSH; virucide;  
 KW fungicide; insecticide; chromosome 5; DNA-binding protein; DNA repair;  
 KW DNA metabolism; DNA replication; cell cycle; homologous recombination;  
 KW pathogen resistance; fungi; virus; nematode; insect; gene therapy;  
 KW genetic manipulation.

XX Zea mays.

OS WO200015816-A2.

PN 23-MAR-2000.

XX 15-SEP-1999; 99WO-US21277.

XX 17-SEP-1998; 98US-0100690.

PR 11-MAR-1999; 99US-0123896.

XX (PION-) PIONEER HI-BRED INT INC.

XX Mahajan P;

XX WPI; 2000-271452/23.

DR N-PSDB; AAZ52289.

XX New maize replication protein A useful for genetic transformation, gene  
 PT targeting in plants and modulating DNA metabolism -

XX Claim 1c; Page 93; 101pp; English.

XX

CC The present sequence is the maize replication protein A (RPA) middle  
 CC subunit homologue-4 (ZMRPAMSH4). RPA is a single-stranded DNA-binding  
 CC protein required for multiple processes in DNA metabolism, like DNA  
 CC replication, repair mechanism (e.g. nucleotide excision and double  
 CC stranded (ds) DNA break repair) and recombination. The gene for the  
 CC ZMRPAMS homologue is mapped to chromosome 5. This sequence has  
 CC fungicide, virucide and insecticidal activity. DNA encoding RPA is  
 CC used for modulating DNA metabolism, influencing cell cycle, enhancing  
 CC homologous recombination and increasing pathogen resistance in plants.  
 CC Pests that can be controlled include fungal pathogens, viruses,  
 CC nematodes and insects. Antisense sequences can be used to block RPA  
 CC expression and promote non-specific recombination events. RPA protein  
 CC can be used to improve genetic manipulation and also in gene therapy.

XX Sequence 273 AA;

Query Match 58.2%; Score 46; DB 21; Length 273;  
 Best Local Similarity 58.8%; Pred. No. 22;  
 Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 SSPSSTKSSPSNVKSAS 17

Db 190 stptslksspapvtsgs 206  
 I:|:| ||||: | | |

Search completed: January 29, 2002, 10:59:18  
 Job time: 2242 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 10:24:14 ; Search time 32.24 seconds  
(without alignments)  
11.866 Million cell updates/sec

Title: US-09-763-397A-22

Perfect score: 79

Sequence: 1 SSPSTKSSPSNVKSAS 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCRUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	62.0	530	1	US-08-187-793-4
2	48	60.8	328	4	Sequence 4, Appli
3	43	54.4	145	3	US-09-300-672-2
4	41	51.9	143	6	US-08-808-599A-41
5	40	50.6	345	1	Patent No. 5260432
6	40	50.6	345	1	Sequence 1, Appli
7	40	50.6	345	3	US-08-843-993-1
8	40	50.6	345	3	Sequence 1, Appli
9	40	50.6	413	2	US-09-059-520A-1
10	40	50.6	413	3	US-08-481-814A-8
11	39	49.4	126	5	Sequence 2, Appli
12	39	49.4	126	5	US-09-265-566-2
13	39	49.4	147	2	Sequence 1, Appli
14	39	49.4	147	2	PCT-US96-03916-15
15	39	49.4	516	2	Sequence 4, Appli
16	39	49.4	516	2	US-08-866-751A-5
17	39	49.4	647	3	US-08-676-166A-2
18	39	49.4	647	3	US-08-483-577A-148
19	39	49.4	647	4	Sequence 2, Appli
20	39	49.4	648	1	US-08-897-438-148
21	39	49.4	648	1	Sequence 1, Appli
22	39	49.4	648	1	US-08-276-151-2
23	39	49.4	648	1	US-08-185-282-1
24	39	49.4	648	1	US-08-185-282-2
25	39	49.4	648	1	US-08-185-282-3
26	39	49.4	648	1	US-08-185-282-4
27	39	49.4	648	1	US-08-185-282-5
28	39	49.4	648	1	US-08-185-282-12
29	39	49.4	648	2	Sequence 12, Appli
30	39	49.4	648	3	US-08-886-751A-6
31	39	49.4	648	3	Sequence 6, Appli
32	39	49.4	648	4	US-09-209-668-13
33	39	49.4	648	4	Sequence 13, Appli
34	39	49.4	648	4	US-08-971-207-1
35	39	49.4	657	3	Sequence 1, Appli
36	39	49.4	657	3	US-08-613-009A-19
37	39	49.4	657	3	Sequence 19, Appli

28	39	49.4	660	1	US-08-487-890A-8	Sequence 8, Appli
29	39	49.4	660	1	US-08-487-890A-10	Sequence 10, Appli
30	39	49.4	660	2	US-08-478-435-8	Sequence 8, Appli
31	39	49.4	660	2	US-08-478-435-10	Sequence 10, Appli
32	39	49.4	660	2	US-08-337-483-8	Sequence 8, Appli
33	39	49.4	660	2	US-08-337-483-10	Sequence 10, Appli
34	39	49.4	660	2	US-08-478-373-8	Sequence 8, Appli
35	39	49.4	660	2	US-08-478-373-10	Sequence 10, Appli
36	39	49.4	660	3	US-08-474-671-8	Sequence 8, Appli
37	39	49.4	660	3	US-08-474-671-10	Sequence 10, Appli
38	39	49.4	660	3	US-08-483-577A-8	Sequence 8, Appli
39	39	49.4	660	3	US-08-483-577A-10	Sequence 10, Appli
40	39	49.4	660	4	US-08-897-438-8	Sequence 8, Appli
41	39	49.4	660	4	US-08-897-438-10	Sequence 10, Appli
42	39	49.4	951	3	US-08-816-346-58	Sequence 58, Appli
43	39	49.4	951	4	US-05-335-411-58	Sequence 58, Appli
44	39	49.4	952	2	US-08-788-674-5	Sequence 5, Appli
45	39	49.4	952	3	US-08-816-346-4	Sequence 4, Appli

#### ALIGNMENTS

RESULT 1  
US-08-187-793-4  
; Sequence 4, Application US/08187793  
; Patent No. 5614371  
; GENERAL INFORMATION:  
; APPLICANT: POSNER, JEROME B.  
; APPLICANT: DARNELL, ROBERT B.  
; APPLICANT: FURNEAUX, HENRY M.  
; TITLE OF INVENTION: ANTIGEN RECOGNIZED BY ANTIBODIES ASSOCIATED  
; TITLE OF INVENTION: WITH PARANEOPlastic OPsOcLONUS AND METHODS OF  
; TITLE OF INVENTION: USE THEREOF  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: COOPER & DUNHAM  
; STREET: 30 ROCKEFELLER PLAZA  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: U.S.A.  
; ZIP: 10112  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PatentIn Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/187,793  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 691,559  
; FILING DATE: 25-APR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WHITE, JOHN P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 39227  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 977-9550  
; TELEFAX: (212) 664-0525  
; TELEX: (212) 422523 COOP UI  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 530 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-187-793-4

Query Match 62.0%; Score 49; DB 1; Length 530;  
Best Local Similarity 58.8%; Pred. No. 5.3;

us-09-763-397a-22.ra1

Mon Feb 4 15:23:44 2002

```

; REFERENCE/DOCKET NUMBER: P-LA 2256
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 145 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-808-599A-41

Query Match          54.4%; Score 43; DB 3; Length 145;
Best Local Similarity 53.3%; Pred. No. 9.9;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 PSTKSSPSNVKSAS 17
Db 1 PSTSTSPSSTSTSS 15

RESULT 4
5260432-3
; Patent No. 5260432
; APPLICANT: TAKAKU, FUMIMARO; ISHIKAWA, TAKASHI; IMAWARI, MICHIO;
; EVANS, RONALD M.; UMESONO, KAZUHIKO
; TITLE OF INVENTION: HUMAN GAMMA RETINOIC ACID RECEPTOR DNA
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/370,407
; FILING DATE: 22-JUN-1989
; SEQ ID NO: 3
; LENGTH: 143
; 5260432-3

Query Match          51.9%; Score 41; DB 6; Length 143;
Best Local Similarity 72.7%; Pred. No. 19;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SSPSTKSSPS 11
Db 129 SSPSSNRSSPA 139

RESULT 5
US-08-843-993-1
; Sequence 1, Application US/08843993
; Patent No. 5739010
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN TRANSCRIPTION FACTOR
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER, READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; APPLICATION NUMBER: US/08/843,993
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:

```

NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0274 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 345 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: SYN00AT01  
CLONE: 727885  
US-08-843-993-1

Query Match 50.6%; Score 40; DB 1; Length 345;  
Best Local Similarity 52.9%; Pred. No. 75;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SSPSTKSSPSNVKSAS 17  
Db 70 SSPSAGNSPSSLKPPS 86

RESULT 6  
US-09-059-520A-1  
; Sequence 1, Application US/09059520A  
; Patent No. 6001971  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; TITLE OF INVENTION: HUMAN TRANSCRIPTION FACTOR  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/059,520A  
; FILING DATE: Filed Herewith  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0274 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 345 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: SYN00AT01  
; CLONE: 727885  
US-09-059-520A-1

Query Match 50.6%; Score 40; DB 3; Length 345;  
Best Local Similarity 52.9%; Pred. No. 75;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SSPSTKSSPSNVKSAS 17  
Db 70 SSPSAGNSPSSLKPPS 86

RESULT 7  
US-09-334-275-1  
; Sequence 1, Application US/09334275  
; Patent No. 6037132  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; TITLE OF INVENTION: HUMAN TRANSCRIPTION FACTOR  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/334,275  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/059,520  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0274 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 345 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: SYN00AT01  
; CLONE: 727885  
US-09-334-275-1

Query Match 50.6%; Score 40; DB 3; Length 345;  
Best Local Similarity 52.9%; Pred. No. 75;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SSPSTKSSPSNVKSAS 17  
Db 70 SSPSAGNSPSSLKPPS 86

RESULT 8  
US-08-481-814A-8  
; Sequence 8, Application US/08481814A  
; Patent No. 5869040  
; GENERAL INFORMATION:  
; APPLICANT: Hsu, Yen-Ming  
; TITLE OF INVENTION: GENE THERAPY METHODS AND COMPOSITIONS  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:

```

; ADDRESSSEE: Biogen, Inc.
; STREET: 14 Cambridge Center
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 02142
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481.814A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Kaplan, Warren A
; REFERENCE/DOCKET NUMBER: A001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-679-2000
; TELEFAX: 617-679-2838
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 413 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-481-814A-8
;
Query Match          50.6%; Score 40; DB 2; Length 413;
Best Local Similarity 58.8%; Pred. No. 92;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY      1 SSPSSTKSSPSNVKSAS 17
Db      309 SSSSSSSSSSSNSNSS 325
        |||||:|||||:|:|

RESULT    9
US-08-836-582-2
; Sequence 2, Application US/08836582
; Patent No. 6045999
; GENERAL INFORMATION:
; APPLICANT: Bernards, Rene
; APPLICANT: Beijersbergen, Roderick L
; TITLE OF INVENTION: Transcription factor E2F-4
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye
; STREET: 1100 No. 6045999th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: US
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,582
; FILING DATE: 10-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/00868
; FILING DATE: 18-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9423049.7
; FILING DATE: 15-NOV-1994
;
Query Match          50.6%; Score 40; DB 2; Length 413;
Best Local Similarity 58.8%; Pred. No. 92;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY      1 SSPSSTKSSPSNVKSAS 17
Db      309 SSSSSSSSSSSNSNSS 325
        |||||:|||||:|:|

ATTORNEY/AGENT INFORMATION:-
NAME: Arthur R. Crawford
REGISTRATION NUMBER: 25,327
REFERENCE/DOCKET NUMBER: 620-18
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 413 amino acids
TYPE: amino acid
MOLECULE TYPE: protein
TOPOLOGY: linear
US-08-836-582-2
;
Query Match          50.6%; Score 40; DB 3; Length 413;
Best Local Similarity 58.8%; Pred. No. 92;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY      1 SSPSSTKSSPSNVKSAS 17
Db      309 SSSSSSSSSSSNSNSS 325
        |||||:|||||:|:|

RESULT   10
US-09-265-566-2
; Sequence 2, Application US/09265566
; Patent No. 6303335
; GENERAL INFORMATION:
; APPLICANT: Bernards, Rene
; APPLICANT: Beijersbergen, Roderick L
; TITLE OF INVENTION: Transcription factor E2F-4
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye
; STREET: 1100 No. 6303335th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: US
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/265,566
; FILING DATE: 10-Mar-1999
; CLASSIFICATION: <Unknown>
; 15-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,582
; FILING DATE: <Unknown>
; APPLICATION NUMBER: GB 9423049.7
; FILING DATE: 15-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Arthur R. Crawford
; REGISTRATION NUMBER: 25,327
; REFERENCE/DOCKET NUMBER: 620-18
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 413 amino acids
; TYPE: amino acid
; MOLECULE TYPE: protein
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-265-566-2
;
Query Match          50.6%; Score 40; DB 4; Length 413;
Best Local Similarity 58.8%; Pred. No. 92;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY      1 SSPSSTKSSPSNVKSAS 17
        |||||:|||||:|:|

```

Db 309 SSSSSSSSSSSNSSSS 325

## RESULT 11

PCT-US96-03916-15  
; Sequence 15, Application PC/TUS9603916  
; GENERAL INFORMATION:  
; APPLICANT: Wild, Martha A.  
; APPLICANT: Cochran, Mark D.  
; TITLE OF INVENTION: RECOMBINANT INFECTIOUS LARYNGOTRACHEITIS VIRUS  
; TITLE OF INVENTION: AND USES THEREOF  
; NUMBER OF SEQUENCES: 72  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/03916  
; FILING DATE: 23-MAR-1995

CLASSIFICATION:  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 08/126,597  
; FILING DATE: 24-SEP-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 39116-A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 278-0400  
; TELEFAX: (212) 391-0525  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 126 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
PCT-US96-03916-15

Query Match 49.4%; Score 39; DB 5; Length 126;

Best Local Similarity 61.5%; Pred. No. 34;

Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 2 SPSSTKSSPSNVK 14

Db 10 SPSDHASSPANAK 22

## RESULT 12

US-08-276-151-4  
; Sequence 4, Application US/08276151  
; Patent No. 5597719  
; GENERAL INFORMATION:  
; APPLICANT: Freed, Ellen  
; APPLICANT: Ruggieri, Rosamaria  
; TITLE OF INVENTION: Interaction of raf-1 and 14-3-3 Proteins  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooley Godward et al.  
; STREET: Five Palo Alto Square  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94036

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/886,751A  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Young, Leona G.  
; REGISTRATION NUMBER: 37,266  
; REFERENCE/DOCKET NUMBER: 25490-0100  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (404) 818-3700  
; TELEFAX: (404) 818-3799  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 217 amino acids

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/276,151  
; FILING DATE: 14-JUL-1994  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Torchia, Ph.D., Timothy E  
; REGISTRATION NUMBER: 36,700  
; REFERENCE/DOCKET NUMBER: ONYX-005/0005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 843-5481  
; TELEFAX: (415) 857-0663  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 147 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; FRAGMENT TYPE: Internal  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
US-08-276-151-4

Query Match 49.4%; Score 39; DB 1; Length 147;  
Best Local Similarity 53.8%; Pred. No. 40;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 SPSSTKSSPSNV 13

Db 103 ASPSALSSPNL 115

## RESULT 13

US-08-886-751A-5  
; Sequence 5, Application US/08886751A  
; Patent No. 5885783  
; GENERAL INFORMATION:  
; APPLICANT: Yoo, Tai-June  
; APPLICANT: Cheng, Kuang-Chuan  
; TITLE OF INVENTION: Autoimmune Inner Ear Disease Antigen and  
; TITLE OF INVENTION: Diagnostic Assay  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Jones & Askew  
; STREET: 191 Peachtree Street, 37th Floor  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: U.S.A.  
; ZIP: 30303

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/886,751A  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Young, Leona G.  
; REGISTRATION NUMBER: 37,266  
; REFERENCE/DOCKET NUMBER: 25490-0100  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (404) 818-3700  
; TELEFAX: (404) 818-3799  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 217 amino acids





**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 10:26:50 ; Search time 34.94 seconds  
(without alignments)  
37.063 Million cell updates/sec

Title: US-09-763-397A-22

Perfect score: 79

Sequence: 1 SSPSSTKSSPSNVKSAS 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_68.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	68.5	86.7	782	2 S27833	rhostry-associated
2	49	62.0	510	2 I38489	onconeural ventral
3	48	60.8	328	2 G84826	hypothetical prote
4	48	60.8	725	2 A41258	a-agglutinin core
5	48	60.8	1446	1 A45344	immediate-early pr
6	47	59.5	497	2 E82537	hypothetical prote
7	47	59.5	515	2 A25048	regulatory protein
8	46	58.2	244	2 T12458	hypothetical prote
9	45	57.0	211	2 D96507	hypothetical prote
10	45	57.0	349	2 T41394	hypothetical serin
11	45	57.0	626	2 JC5405	brain and muscle A
12	45	57.0	626	2 JE0270	Antt-like PAS prot
13	45	57.0	2559	2 T09144	probable guanine n
14	44	55.7	214	2 T51657	myb-related transc
15	44	55.7	229	2 T04764	hypothetical prote
16	44	55.7	267	2 H86320	probable MYB47 tra
17	44	55.7	892	2 A41697	nitrate assimilati
18	44	55.7	928	2 T24868	hypothetical prote
19	44	55.7	1194	2 E96624	hypothetical prote
20	43	54.4	230	2 D86352	protein T26F17.14
21	43	54.4	230	2 T39598	hypothetical serin
22	43	54.4	327	2 G01037	T-cell activation
23	43	54.4	388	2 S04783	protoporphyrinoid
24	43	54.4	402	2 E86185	hypothetical prote
25	43	54.4	448	2 B56558	retinoic acid rece
26	43	54.4	541	2 D84681	hypothetical prote
27	43	54.4	542	2 T06728	pectate lyase [EC
28	43	54.4	701	2 T52384	hypothetical prote
29	43	54.4	969	2 F71418	hypothetical prote

hypothetical prote  
unconventional myo  
vitellogenin A2 pr  
hypothetical prote  
immediate-early pr  
probable lysophosp  
probable leucine-r  
retinoic acid rece  
retinoic acid rece  
hypothetical prote  
probable membrane  
serine-rich protei  
related to BCS1 pr  
hypothetical prote

#### ALIGNMENTS

##### RESULT 1

S27833

rhostry-associated protein 1 precursor - malaria parasite (Plasmodium falciparum)  
N:Alternate names: protective antigen  
C:Species: Plasmodium falciparum  
C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 09-Jun-2000  
C:Accession: A45514; S27833

R:Ridley, R.G.; Takacs, B.; Lahm, H.W.; Delves, C.J.; Goman, M.; Certa, U.; Matile, H  
Mol. Biochem. Parasitol. 41, 125-134, 1990  
A:Title: Characterisation and sequence of a protective rhostry antigen from Plasmodiu  
A:Reference number: A45514; MUID:90348711  
A:Accession: A45514

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-782 <RI2>

A:Cross-references: GB:M32853; NID:g160656; PID:g160657

C:Superfamily: Plasmodium falciparum rhostry-associated protein 1

Query Match 86.7%; Score 68.5; DB 2; Length 782;

Best Local Similarity 94.4%; Pred. No. 0.014;

Matches 17; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 SSPSSTK-SSPSNVKSAS 17

Db 145 SSPSSTKSSPSNVKSAS 162

##### RESULT 2

I38489

onconeural ventral antigen-1 - human

C:Species: Homo sapiens (man)

C:Date: 06-Sep-1996 #sequence\_revision 06-Sep-1996 #text\_change 05-Nov-1999

C:Accession: I38489

R:Buckanovich, R.J.; Posner, J.B.; Darnell, R.B.  
Neuron 11, 657-672, 1993

A:Title: Nova, the Paraneoplastic Ri Antigen, is Homologous to an RNA-Binding Protein

A:Reference number: I38489; MUID:94000830

A:Accession: I38489

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-510 <RES>

A:Cross-references: EMBL:U04840; NID:g440877; PIDN:AAAI6022.1; PID:g440878

C:Genetics:

A:Gene: Nova-1

Query Match

Best Local Similarity 62.0%; Score 49; DB 2; Length 510;

Matches 10; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 SSPSSTKSSPSNVKSAS 17



## RESULT 7

A25048  
regulatory protein STE7 - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein D1525; protein YDL159w  
C:Species: Saccharomyces cerevisiae  
C>Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 24-Sep-1999  
C:Accession: A25048; S61054; S67711; S67707  
R:Teague, M.A.; Chaleff, D.T.; Errede, B.  
Proc. Natl. Acad. Sci. U.S.A. 83, 7371-7375, 1986  
A:Title: Nucleotide sequence of the yeast regulatory gene STE7 predicts a protein homologous to the protein kinase homology domain of the yeast regulatory gene STE7  
A:Reference number: A25048; MUID:87016949  
A:Accession: A25048  
A:Molecule type: DNA  
A:Residues: 1-515 <TEA>  
A:Cross-references: EMBL:M14097; NID:g172761; PIDN:AAA35118.1; PID:g172762  
R:Pohl, T.M.  
submitted to the EMBL Data Library, November 1995  
A:Reference number: S61010  
A:Accession: S61054  
A:Molecule type: DNA  
A:Residues: 1-515 <POH>  
A:Cross-references: EMBL:Z67750; NID:g1061256; PIDN:CAA91587.1; PID:g1061280  
R:Pohl, T.M.  
submitted to the Protein Sequence Database, July 1996  
A:Reference number: S67708  
A:Accession: S67711  
A:Molecule type: DNA  
A:Residues: 1-515 <POW>  
A:Cross-references: EMBL:Z74207; NID:g1431250; PIDN:CAA98732.1; PID:e253067; PID:g143125  
A:Experimental source: strain S288C  
R:Perea, J.; Blugeon, C.; Delaveau, T.; Jacq, C.  
submitted to the Protein Sequence Database, July 1996  
A:Reference number: S67693  
A:Accession: S67707  
A:Molecule type: DNA  
A:Residues: 71-515 <PER>  
A:Cross-references: EMBL:Z74207; MIPS:YDL159w  
A:Experimental source: strain S288C  
C:Genetics:  
A:Gene: SGD:STE7  
A:Cross-references: SGD:S0002318; MIPS:YDL159w  
A:Map position: 4L  
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homology  
C:Keywords: ATP; phosphoprotein; serine/threonine-specific protein kinase  
F:189-466/Domain: protein kinase homology <KIN>  
F:197-205/Region: protein kinase ATP-binding motif

Query Match 59.5%; Score 47; DB 2; Length 515;  
Best Local Similarity 56.2%; Pred. No. 15;  
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 SPSSTKSSPSNVKSAS 17  
|||||I:|:|:|  
Db 130 SPSSTNSTPSTIQGLS 145

## RESULT 8

T12458  
hypothetical protein DKFp56400823.1 - human  
C:Species: Homo sapiens (man)  
C>Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 23-Jul-1999  
C:Accession: T12458  
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, June 1999  
A:Reference number: Z17524  
A:Accession: T12458  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-244 <WAM>  
A:Cross-references: EMBL:AL080121

A:Experimental source: fetal brain; clone DKFp56400823  
C:Genetics:  
A:Note: DKFp56400823.1

Query Match 58.2%; Score 46; DB 2; Length 244;  
Best Local Similarity 64.7%; Pred. No. 9.8;  
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 SSPSSTKSSPSNVKSAS 17  
|||||I:|:|:|  
Db 156 SSPSSLSTSPPEVFSAS 172

## RESULT 9

D96507  
hypothetical protein T12C22.10 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: D96507  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
ansen, N.P.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marzia  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: D96507  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-211 <STO>  
A:Cross-references: GB:AE005173; NID:g8655993; PIDN:AAF78266.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: T12C22.10  
A:Map position: 1

Query Match 57.0%; Score 45; DB 2; Length 211;  
Best Local Similarity 58.8%; Pred. No. 12;  
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 SSPSSTKSSPSNVKSAS 17  
|||||I:|:|:|  
Db 13 SSPSSSSSSSTSSSS 29

## RESULT 10

T41394  
hypothetical serine-rich protein - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T41394  
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Murphy, L.; Harris, D.  
submitted to the EMBL Data Library, May 1998  
A:Reference number: Z21991  
A:Accession: T41394  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-349 <WOO>  
A:Cross-references: EMBL:AL023704; PIDN:CAA19262.1; GSPDB:GN00068; SPDB:SPCC553.10  
A:Experimental source: strain 972h-; cosmid C553  
C:Genetics:  
A:Gene: SPDB:SPCC553.10  
A:Map position: 3

Query Match 57.0%; Score 45; DB 2; Length 349;  
Best Local Similarity 64.7%; Pred. No. 20;  
Matches 11; Conservative 3; Mismatches 3; Indels 0; Gaps 0;





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 11:13:51 ; Search time 20.36 seconds  
(without alignments)  
30.614 Million cell updates/sec

Title: US-09-763-397A-22

Perfect score: 79

Sequence: 1 SSPSSTKSSPSNVKSAS 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	49	62.0	510	1 NOAL_HUMAN	P51513 homo sapien
2	48	60.8	725	1 AGAL_YEAST	P32323 saccharomyc
3	48	60.8	1446	1 IE18_PRIVK	P33479 pseudorabie
4	47	59.5	515	1 STE7_YEAST	P06784 saccharomyc
5	46	58.2	710	1 IRAL_MOUSE	O62406 mus musculu
6	45	57.0	583	1 BMAL_HUMAN	O00327 homo sapien
7	44	55.7	234	1 RACD_DICDI	P34150 dictyostell
8	44	55.7	892	1 NIRA_EMENI	P28348 emericeella
9	43	54.4	388	1 PORA_HORVU	P13653 hordeum vul
10	43	54.4	583	1 API_KLULA	P56095 kluyveromyc
11	43	54.4	681	1 BRC2_HALRO	O01409 halocynthia
12	43	54.4	1807	1 VTA2_XENLA	P18709 xenopus lae
13	42.5	53.8	1461	1 IE18_PRIVF	P11675 pseudorabie
14	42	53.2	458	1 RRA_XENLA	P51126 xenopus lae
15	42	53.2	503	1 YN23_YEAST	P53832 saccharomyc
16	42	53.2	507	1 YG46_YEAST	P53301 saccharomyc
17	42	53.2	796	1 YS8A_CAEEL	O09625 caenorhabdi
18	42	53.2	845	1 CLPC_CHLPN	P02846 chlamydia p
19	42	53.2	1214	1 SIPI_MOUSE	O60315 homo sapien
20	42	53.2	1215	1 SIPI_MOUSE	O9097 mus musculu
21	42	53.2	1475	1 N153_HUMAN	P49790 homo sapien
22	42	53.2	1627	1 TP2B_CHICK	O42131 gallus gall
23	42	53.2	1850	1 VIT2_CHICK	P02845 gallus gall
24	41	51.9	485	1 Y136_TREPA	O83172 treponema p
25	41	51.9	561	1 CP51_USTMA	P49602 ustilago ma
26	41	51.9	1071	1 SEFI_KLULA	P87164 kluyveromyc
27	41	51.9	1091	1 NCAL_CHICK	P13590 gallus gall
28	41	51.9	1147	1 AC15_HUMAN	P35251 homo sapien
29	41	51.9	1597	1 RLRI_YEAST	P53552 saccharomyc
30	41	51.9	1992	1 TR12_HUMAN	O14669 homo sapien
31	40.5	51.3	483	1 ODP2_SCHPO	O59816 schizosacch
32	40	50.6	161	1 L9KD_MYCAV	P46733 mycobacteri
33	40	50.6	304	1 CAH5_RAT	P43165 rattus norv

ALIGNMENTS

RESULT 1

ID	NOAL_HUMAN	STANDARD;	PRT;	510 AA.
AC	P51513;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	ONCONEURAL VENTRAL ANTIGEN-1 (NOVA-1) (PARANEOPLASTIC RI ANTIGEN)			
DE	(VENTRAL NEURON-SPECIFIC PROTEIN 1).			
GN	NOVA1			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Cerebellum, and Hippocampus;			
RX	MEDLINE=94000830; PubMed=8398153;			
RA	Buckanovich R.J., Posner J.B., Darnell R.B.;			
RA	"Nova, the paraneoplastic RI antigen, is homologous to an RNA-binding			
RT	protein and is specifically expressed in the developing motor			
RT	system."			
RL	Neuron 11:657-672(1993).			
RN	[2]			
RC	SEQUENCE OF 1-34 FROM N.A.			
TISSUE	Fetal brain;			
RA	Dmitrenko V.V., Garifulin O.M., Shostak K.A., Smikodub A.I.,			
RA	Kavsan V.M.;			
Submitted	(APR-1996) to the EMBL/GenBank/DBJ databases.			
CC	-1- FUNCTION: MAY REGULATE RNA SPLICING OR METABOLISM IN A SPECIFIC			
CC	SUBSET OF DEVELOPING NEURONS.			
CC	-1- SUBCELLULAR LOCATION: NUCLEAR.			
CC	-1- TISSUE SPECIFICITY: BRAIN.			
CC	-1- DISEASE: ANTIBODIES ARE SEEN AGAINST THIS ANTIGEN IN THE PATIENTS			
CC	SUFFERING WITH PARANEOPLASTIC OPSOCLONUS-ATAXIA (POA). POA IS A			
CC	DISORDER IN WHICH ABNORMAL MOTOR CONTROL OF THE EYES, TRUNK AND			
CC	LIMBS DEVELOP IN WOMEN WITH BREAST OR SMALL LUNG CANCER. POA			
CC	PATIENTS SUFFER FROM OPSOCLONUS, A CHAOTIC EYE MOVEMENT DISORDER			
CC	ATTRIBUTABLE TO A LACK OF INHIBITION OF BURST NEURONS IN THE BRAIN			
CC	STEM; MYOCLONUS, A DISORDER PERHAPS ATTRIBUTING TO A FAILURE OF			
CC	INHIBITORY CONTROL OVER SPINAL MOTOR NEURONS; AND TRUNCAL ATAXIA,			
CC	REFERABLE TO CEREBELLAR DYSFUNCTION; THERE ARE NO CORTICAL OR			
CC	SENSORY DEFICITS.			
CC	-1- SIMILARITY: CONTAINS 3 KH DOMAINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; U04840; AAA16022.1; -			
DR	EMBL; Z70771; CAA94810.1; -			
DR	MIM; 602157; -			

34	40	50.6	345	1	KLF3_HUMAN	P57682 homo sapien
35	40	50.6	355	1	YP11_YEAST	P48559 saccharomyc
36	40	50.6	372	1	COQ2_YEAST	P32378 saccharomyc
37	40	50.6	376	1	MID2_YEAST	P36027 saccharomyc
38	40	50.6	413	1	E2F4_HUMAN	Q16254 homo sapien
39	40	50.6	417	1	MAC1_YEAST	P35192 saccharomyc
40	40	50.6	702	1	YAI4_YEAST	P27625 saccharomyc
41	40	50.6	709	1	SYR_CAEEL	Q19825 caenorhabdi
42	40	50.6	743	1	YCR6_YEAST	P25353 saccharomyc
43	40	50.6	754	1	UBP9_YEAST	P39967 saccharomyc
44	40	50.6	823	1	YRR2_CAEEL	Q09345 caenorhabdi
45	40	50.6	888	1	STU2_YEAST	P46675 saccharomyc

DR InterPro: IPR000958; KH.  
DR Pfam: PF00013; KH-domain; 3.  
DR SMART: SM00322; KH; 3.  
DR PROSITE: PS50084; KH\_TYPE\_1; 3.  
KW Nuclear protein; RNA-binding; Repeat; Alternative splicing; Antigen.  
FT DOMAIN 52 87  
FT DOMAIN 177 212  
FT DOMAIN 427 462  
FT VARSPLIC 88 90 MISSING (IN TUMOR ISOFORM).  
FT VARSPLIC 133 176 MISSING (IN ISOFORM 3).  
FT VARSPLIC 177 184 VKIVPNS -> KHNISWIS (IN TUMOR ISOFORM).  
FT VARSPLIC 165 510 MISSING (IN TUMOR ISOFORM).  
SQ SEQUENCE 510 AA; 52055 MW; 7B9BA8B67CAB6984 CRC64;

Query Match 62.0%; Score 49; DB 1; Length 510;  
Best Local Similarity 58.8%; Pred. No. 3.9;  
Matches 10; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 SSPSTKSSPSNVKAS 17  
|||||:||||: :  
Db 156 SPPTTKSSPSDPMTS 172

RESULT 2  
AGAL\_YEAST  
ID AGAL\_YEAST STANDARD; PRT; 725 AA.  
AC P32323;  
DT 01-OCT-1993 (Rel. 27, Created)  
DE 01-OCT-1993 (Rel. 27, Last sequence update)  
DE 01-OCT-1996 (Rel. 34, Last annotation update)  
DE A-AGGLUTININ ATTACHMENT SUBUNIT PRECURSOR.  
GN AGAL OR YNR044W OR N3431.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Pohl T.M.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: MEDIATES THE CELL SURFACE ATTACHMENT OF THE A-AGGLUTININ  
CC SUBUNIT. S. CEREVISIAE A AND ALPHA CELLS EXPRESS THE COMPLEMENTARY  
CC CELL SURFACE GLYCOPROTEINS A-AGGLUTININ AND ALPHA-AGGLUTININ,  
CC RESPECTIVELY, WHICH INTERACT WITH ONE ANOTHER TO PROMOTE CELLULAR  
CC AGGREGATION DURING MATING.  
CC -!- SUBUNIT: CONTAINS AT LEAST A BINDING SUBUNIT DISULFIDE-LINKED TO  
CC A CORE SUBUNIT.  
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR  
CC AND/OR BY CARBOHYDRATE-MEDIATED COVALENT CROSS-LINKS (POSSIBLE).  
CC -!- PTM: EXTENSIVELY O-GLYCOSYLATED.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M60590; AAA34382.1; -;  
DR EMBL; Z71659; CAA96325.1; -;  
DR PIR; S17031; S17031.  
DR PIR; A41258; A41258.  
DR SGD; S0005327; AGAL.  
KW Glycoprotein; Cell adhesion; Signal; GPI-anchor; Repeat;

KW Pheromone response.  
FT SIGNAL 1 22  
FT CHAIN 23 725  
FT DOMAIN 53 493  
FT REPEAT 53 149  
FT REPEAT 395 493  
FT DOMAIN 182 307  
FT REPEAT 182 188  
FT REPEAT 189 195  
FT REPEAT 196 202  
FT REPEAT 203 209  
FT REPEAT 210 216  
FT REPEAT 217 223  
FT REPEAT 224 230  
FT REPEAT 231 237  
FT REPEAT 238 244  
FT REPEAT 245 251  
FT REPEAT 252 258  
FT REPEAT 259 265  
FT REPEAT 266 272  
FT REPEAT 273 279  
FT REPEAT 280 286  
FT REPEAT 287 293  
FT REPEAT 294 300  
FT REPEAT 301 307  
SQ SEQUENCE 725 AA; 73353 MW; 70420C853B0B01F8 CRC64;

Query Match 60.8%; Score 48; DB 1; Length 725;  
Best Local Similarity 52.9%; Pred. No. 7.9;  
Matches 9; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 1 SSPSTKSSPSNVKAS 17  
|||||:||||: :  
Db 191 TSPSTSTSPSTSTSS 207

RESULT 3  
IE18\_PVKA  
ID IE18\_PVKA STANDARD; PRT; 1446 AA.  
AC P33479; 1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-FEB-1994 (Rel. 28, Last annotation update)  
DE IMMEDIATE-EARLY PROTEIN IE180.  
GN IE.  
OS Pseudorabies virus (strain Kaplan) (PRV).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Varicellovirus.  
OX NCBI\_TaxID=33703;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91021039; PubMed=2171211;  
RA Vicek C., Kozmik Z., Paces V., Schirm S., Schwyzer M.;  
RT "Pseudorabies virus immediate-early gene overlaps with an oppositely  
RT oriented open reading frame: characterization of their promoter and  
RT enhancer regions."  
RL Virology 179:365-377(1990).  
CC -!- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE  
CC OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING  
CC OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS.  
CC -!- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.  
CC -!- PTM: A LONG STRETCH OF SERINE RESIDUES MAY BE A MAJOR SITE OF  
CC PHOSPHORYLATION.  
CC -!- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).



```
CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL; M34651; AAA47470.1; -.
CC PIR; A45344; A45344.
CC Early protein; Transcription regulation; Trans-acting factor;
KW DNA-binding; Phosphorylation; Nuclear protein.
FT DOMAIN 347 354 POLY-SER.
FT DOMAIN 379 397 POLY-SER.
SQ SEQUENCE 1446 AA; 148640 MW; 81F43A3DE3DDA068 CRC64;

Query Match 60.8%; Score 48; DB 1; Length 1446;
Best Local Similarity 64.7%; Pred. No. 16;
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 SSPSTKSSPSNVKKSAS 17
Db 842 SSPASTKSSSSSTKSSSS 858
|||||
| | | | |

RESULT 4
STE7_YEAST STANDARD; PRT; 515 AA.
AC P06784;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SERINE/THREONINE PROTEIN KINASE STE7 (EC 2.7.1.-).
GN STE7 OR YDL159W OR D1525.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87016949; PubMed=3532111;
RA Teague M.A., Chaleff D.T., Errede B.;
RT "Nucleotide sequence of the yeast regulatory gene STE7 predicts a
RT protein homologous to protein kinases."
RL Proc. Natl. Acad. Sci. U.S.A. 83:7371-7375(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RA Pohl T.M.;
RN Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 70-515 FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=97127830; PubMed=8972581;
RA Delaveau T., Blugeon C., Jacq C., Perea J.;
RT "Analysis of a 23 kb region on the left arm of yeast chromosome IV."
RL Yeast 12:1587-1592(1996).
RN [4]
RP POSSIBLE FUNCTION.
RX MEDLINE=92331935; PubMed=1628833;
RA Cairns B.R., Ramer S.W., Kornberg K.D.;
RT "Order of action of components in the yeast pheromone response
RT pathway revealed with a dominant allele of the STE11 kinase and the
RT multiple phosphorylation of the STE7 kinase."
RL Genes Dev. 6:1305-1318(1992).
RN [5]
RP PHOSPHORYLATION SITES SER-369 AND THR-373, AND MUTAGENESIS.
RX MEDLINE=94178255; PubMed=8131746;
RA Zheng C.-F., Guan K.-L.;
RT "Activation of WEK family kinases requires phosphorylation of two
RT conserved Ser/Thr residues."
RL EMBO J. 13:1123-1131(1994).
CC -1- FUNCTION: SERINE/THREONINE PROTEIN KINASE REQUIRED FOR CELL-TYPE-
CC SPECIFIC TRANSCRIPTION AND SIGNAL TRANSDUCTION IN YEAST. IT
CC IS THOUGHT THAT IT IS PHOSPHORYLATED BY THE STE11 PROTEIN KINASE
CC AND THAT IT CAN PHOSPHORYLATE THE FUS3 AND OR KSS1 KINASES.
CC -1- ENZYME REGULATION: PHOSPHORYLATED AT MULTIPLE SITES IN RESPONSE
CC TO PHEROMONE.
```

```
CC -1- PATHWAY: RESPONSE TO PHEROMONE-INDUCED SIGNAL.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC MAP KINASE KINASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL; M14097; AAA35118.1; -.
CC EMBL; Z67750; CAA91587.1; -.
CC EMBL; Z74207; CAA98732.1; -.
CC EMBL; X97751; CAA66332.1; -.
CC PIR; A25048; A25048.
CC SGD; S0002318; STE7.
CC DR InterPro; IPR000719; Euk_pkinase.
CC DR InterPro; IPR002290; Ser_thr_kin_actsite.
CC DR Pfam; PF00069; pkinase; 1.
CC DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
CC DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC DR Transferase; Serine/threonine-protein kinase; ATP-binding;
KW Phosphorylation; Pheromone response.
FT DOMAIN 191 466 PROTEIN KINASE.
FT NP_BIND 197 205 ATP (BY SIMILARITY).
FT BINDING 220 220 ATP (BY SIMILARITY).
FT ACT_SITE 331 331 BY SIMILARITY.
FT MOD_RES 359 359 PHOSPHORYLATION.
FT MOD_RES 363 363 PHOSPHORYLATION.
FT MUTAGEN 353 353 S->A: NO LOSS OF ACTIVITY.
FT MUTAGEN 359 359 S->A: INACTIVATION.
FT MUTAGEN 363 363 T->A: INACTIVATION.
SQ SEQUENCE 515 AA; 57709 MW; C8A75899CFBE8BDE CRC64;

Query Match 59.5%; Score 47; DB 1; Length 515;
Best Local Similarity 56.2%; Pred. No. 7.7;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 SPSTKSSPSNVKKSAS 17
Db 130 SPSTNSTPSTIQGLS 145
|||||
| | | | |

RESULT 5
IRAL_MOUSE STANDARD; PRT; 710 AA.
AC O62406;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE INTERLEUKIN-1 RECEPTOR-ASSOCIATED KINASE 1 (EC 2.7.1.-) (IRAK-1)
DE (IRAK) (PELLE-LIKE PROTEIN KINASE) (MPLK).
GN IRAK1 OR ILIRAK
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Kopp E.B., Ghosh S.;
RT "Cloning of mouse IRAK."
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Harrington M.A.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 34-710 FROM N.A.
RC TISSUE=Embryo;
```

RX MEDLINE-96279287; PubMed-8663605;  
RA Trofimova M., Sprengle A.B., Green M., Sturgill T.W., Goebel M.G.,  
RA Harrington M.A.;  
RT "developmental and tissue-specific expression of mouse pelle-like  
RT protein kinase.";  
RL J. Biol. Chem. 271:17609-17612(1996).  
CC -1- FUNCTION: INVOLVED IN IL-1 PATHWAY. THIS KINASE ASSOCIATES WITH  
CC THE IL-1 RECEPTOR IL1-R-1. THIS ASSOCIATION IS RAPID AND IL-1  
CC DEPENDENT (BY SIMILARITY).  
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN LIVER, FOLLOWED BY KIDNEY  
CC AND SKELETAL MUSCLE.  
CC -1- DEVELOPMENTAL STAGE: EXPRESSED FROM E11 DAY TO E18 DAY.  
CC -1- PTM: AUTOPHOSPHORYLATED. AN EXTENSIVE PHOSPHORYLATION OF IRAK  
CC OCCURS AFTER ITS ASSOCIATION WITH IL1-R-1. THIS STEP COULD BE  
CC LINKED TO THE ACTIVATION OF THE KINASE (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC PELLE SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U56773; AAC52694.2; -;  
DR EMBL; AF103876; AAD13224.1; ALT\_INIT.  
DR MGI; 107420; Il1rak.  
DR InterPro: IPR000719; Euk\_pkinase.  
DR InterPro: IPR002290; Ser\_thr\_kin\_actsite.  
DR Pfam; PF00069; pkinase; 1.  
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
DR PROSITE; PS00108; PROTEIN KINASE ST; 1.  
DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.  
KW Transferase; Serine/threonine-protein kinase; ATP-binding.  
FT DOMAIN 212 521 PROTEIN KINASE  
FT NP\_BIND 218 226 ATP (BY SIMILARITY).  
FT BINDING 239 239 ATP (BY SIMILARITY).  
FT ACT\_SITE 338 338 BY SIMILARITY.  
SQ SEQUENCE 710 AA; 77269 MW; 8A501F002CD3EBD2 CRC64;  
  
Query Match 58.2%; Score 46; DB 1; Length 710;  
Best Local Similarity 75.0%; Pred. NO. 15;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 SPSSTKSPSPN 12  
Db 174 SAPSSTKSPSPS 185  
|:|||||:  
  
RESULT 6  
ID BMAL\_HUMAN STANDARD; PRT; 583 AA.  
AC O00327; O00313; O00314; O00315; O00316; O00317; Q99631; Q99649;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 13-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE BMAL1 PROTEIN (BRAIN AND MUSCLE ARNT-LIKE 1) (MEMBER OF PAS PROTEIN 3)  
DE (MOP3) (BHLH-PAS PROTEIN JAP3).  
GN BMAL1 OR ARNTL  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_Taxid=9606;  
[1]  
RN SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING (ISOFORMS BMAL1 A-F).  
RC TISSUE=Brain;  
RX MEDLINE-97289529; PubMed-9144434;  
RA Ikeda M., Nomura M.;  
RT "cDNA cloning and tissue-specific expression of a novel basic  
RT helix-loop-helix/PAS protein (BMAL1) and identification of

RT alternatively spliced variants with alternative translation  
RT initiation site usage.";  
RL Biochem. Biophys. Res. Commun. 233:258-264(1997).  
RN [2].  
RP SEQUENCE FROM N.A. (ISOFORM MOP3).  
RC TISSUE=Fetal brain;  
RX MEDLINE-97236817; PubMed-9079689;  
RA Hogenesch J.B., Chan W.K., Jackiw V.H., Brown R.C., Gu Y.-Z.,  
RA Pray-Grant M., Bradfield C.A.;  
RT "Characterization of a subset of the basic-helix-loop-helix-PAS  
RT superfamily that interacts with components of the dioxin signaling  
RT pathway.";  
RL J. Biol. Chem. 272:8581-8593(1997).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM BMAL1B).  
RA Tian H., Russell D.W., McKnight S.L.;  
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP INTERACTION WITH CLOCK.  
RX MEDLINE-98279137; PubMed-9616112;  
RA Gekakis N., Staknis D., Nguyen H.B., Davis F.C., Wilsbacher L.D.,  
RA King D.P., Takahashi J.S., Weitz C.J.;  
RT "Role of the clock protein in the mammalian circadian mechanism.";  
RL Science 280:1564-1569(1998).  
CC -1- FUNCTION: CLOCK-BMAL1 HETERODIMERS BIND TO AN E-BOX ELEMENT (3'-  
CACGTG-5'), THEREBY ACTIVATING TRANSCRIPTION OF PER1, AND POSSIBLY  
CC OF OTHER CIRCADIAN CLOCK PROTEINS.  
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
CC BHLH PROTEIN. FORMS AN HETERODIMER WITH CLOCK. INTERACTS WITH  
CC HSP90; WITH AHR IN VITRO, BUT NOT IN VIVO.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 7 ISOFORMS; BMALIA (SHOWN HERE),  
CC BMAL1B/JAP3, BMAL1C, BMAL1D, BMAL1E, BMAL1F AND MOP3; ARE PRODUCED  
CC BY ALTERNATIVE SPLICING.  
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE ADULT BRAIN, SKELETAL  
CC MUSCLE AND HEART.  
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
CC TRANSCRIPTION FACTORS. HOMOLOG OF DROSOPHILA CYCLE PROTEIN.  
CC -1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D89722; BAA19968.1; -;  
DR EMBL; AB000812; BAA19935.1; -;  
DR EMBL; AB000813; BAA19936.1; -;  
DR EMBL; AB000814; BAA19937.1; -;  
DR EMBL; AB000815; BAA19938.1; -;  
DR EMBL; AB000816; BAA19939.1; -;  
DR EMBL; U51627; AAC51213.1; -;  
DR EMBL; U60415; AAB37248.1; -;  
DR MIM; 602550;  
DR InterPro: IPR003015; HLH\_Myc.  
DR InterPro: IPR001092; HLH\_dim.  
DR InterPro: IPR001067; Nucleoslocatr.  
DR InterPro: IPR001610; PAS.  
DR InterPro: IPR000014; PAC.  
DR Pfam; PF00010; HLH; 1.  
DR Pfam; PF00785; PAC; 1.  
DR Pfam; PF00989; PAS; 2.  
DR PRINTS; PR00785; NCTRNSLOCATR.  
DR SMART; SM00353; HLH; 1.  
DR SMART; SM00086; PAC; 1.  
DR SMART; SM00091; PAS; 2.  
DR PROSITE; PS00038; HELIX LOOP HELIX; 1.  
KW Repeat; DNA-binding; Nuclear protein; Transcription regulation;  
KW Alternative splicing.  
FT DNA\_BIND 30 42 BASIC DOMAIN.

```

RESULT 7
RACD_DICDI
ID RACD_DICDI STANDARD; PRRT; 254 AA.
AC F34150; Q9GFS8;
DT 01-FEB-1994 (Rel. 28, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RAS-RELATED PROTEIN RACD.
RACD.
GN RACD.
OS Dictyostellium discoideum (Slime mold).
OC Eukaryota; Mycotozoa; Dictyostellida; Dictyostellium.
NCBI_TaxID=44689;
[1]
RN RP SEQUENCE FROM N.A.
RX STRAIN=AX4;
RC MEDLINE=21127961; PubMed=11222756;
RA Rivero F., Dislich H., Glockner G., Noegel A.A.;
RT "The Dictyostellium discoideum family of Rho-related proteins.";
RL Nucleic Acids Res. 29:1068-1079(2001).
[2]
RN RP SEQUENCE OF 8-194 FROM N.A.
RX STRAIN=AX3;
RC MEDLINE=941124042; PubMed=8294042;
RA Bush J.M. IV, Franek K., Cardelli J.A.;
RT "Cloning and characterization of seven novel Dictyostellium discoideum
rac-related genes belonging to the rho family of GTPases.";
RL Gene 136:61-68(1993).
CC CC -1- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RHO FAMILY.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way

```

DR InterPro; IPR001138; Zn2\_Cy6\_fungal.  
DR Pfam; PF00172; Zn\_clus; 1.

```
DR SMART: SM00066; GAL4; 1.
DR PROSITE: PS00463; ZN2_CV6_FUNGAL_1; 1.
DR PROSITE: PS00464; ZN2_CV6_FUNGAL_2; 1.
KW Transcription regulation; Activator; DNA-binding; Nuclear protein;
KW Zinc: Metal-binding; Nitrate assimilation.
FT DNA_BIND 42 70 ZN(2)-CYS(6); FUNGAL-TYPE.
FT DOMAIN 109 128 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 195 210 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 407 428 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 647 713 PRO-RICH.
FT DOMAIN 842 887 PRO-RICH.
SQ SEQUENCE 892 AA; 99448 MW; 65FD105203A53F59 CRC64;

Query Match 55.7%; Score 44; DB 1; Length 892;
Best Local Similarity 52.9%; Pred. No. 37;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 SSPSSTKSSPSNVKAS 17
Db 652 ASPSTSDSPSPSVSSQS 668

RESULT 9
PORA_HORVU STANDARD; PRT; 388 AA.
AC P13653;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTOCHLOROPHYLLIDE REDUCTASE A, CHLOROPLAST PRECURSOR (EC 1.3.1.33)
DE (PCR A) (NADPH-PROTOCHLOROPHYLLIDE OXIDOREDUCTASE A) (POR A).
GN PORA.
OS Hordeum vulgare (barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC STRAIN=CV. CARINA; TISSUE=leaf;
RX MEDLINE=89364719; PubMed=2671659;
RA Schuelz R., Steinmuller K., Klaas M., Forreiter C., Rasmussen S.,
RA Hillier C., Apel K.;
RT "Nucleotide sequence of a cDNA coding for the NADPH-
RT protochlorophyllide oxidoreductase (PCR) of barley (Hordeum vulgare
RT L.) and its expression in Escherichia coli.";
RL Mol. Gen. Genet. 217:355-361(1989).
CC -1- FUNCTION: PHOTOTRANSFORMATION OF PROTOCHLOROPHYLLIDE (PCHLIDE) TO
CC CHLOROPHYLLIDE (CHLIDE).
CC -1- CATALYTIC ACTIVITY: CHLOROXYLLIDE A + NADP(+) =
CC PROTOCHLOROPHYLLIDE + NADPH.
CC -1- PATHWAY: CHLOROXYLL BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
CC -1- DEVELOPMENTAL STAGE: ACTIVE ONLY TRANSIENTLY IN ETIOLATED
CC SEEDLINGS AT THE BEGINNING OF ILLUMINATION.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY. FOR SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X15869; CAA33879.1; -.
CC PIR: S04783; S04783.
CC InterPro: IPR002198; ADH_short.
CC Pfam: PF00106; adh_short_1.
KW Photosynthesis; Chlorophyll biosynthesis; Oxidoreductase; NADP;
KW Chloroplast; transit peptide; Multigene family.
```

```
FT TRANSIT 1 74 CHLOROPLAST.
FT CHAIN 75 388
SQ SEQUENCE 388 AA; 41181 MW; EB03EF153D96C129 CRC64;

Query Match 54.4%; Score 43; DB 1; Length 388;
Best Local Similarity 60.0%; Pred. No. 22;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 SSPSSTKSSPSNVKAS 15
Db 57 TSPGSTASSPSGKKT 71

RESULT 10
API_KLULA STANDARD; PRT; 583 AA.
AC P56095;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE AP-1-LIKE TRANSCRIPTION FACTOR.
GN VAP1.
OS Kluyveromyces lactis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98102743; PubMed=9439570;
RA Billard P., Dumond H., Bolotin-Fukuhara M.;
RT "Characterization of an AP-1-like transcription factor that mediates
RT an oxidative stress response in Kluyveromyces lactis.";
RL Mol. Gen. Genet. 257:62-70(1997).
CC -1- FUNCTION: MEDIATES OXIDATIVE STRESS RESPONSE. INVOLVED IN BOTH THE
CC OXIDATIVE AND CADMIUM RESPONSE PATHWAYS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: TO OTHER BZIP PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF006499; AAC39320.1; -.
CC HSP: P03069; 2DGC.
CC InterPro: IPR001871; bZIP.
CC Pfam: PF00170; bZIP; 1.
CC SMART: SM00338; BRLZ; 1.
CC PROSITE: PS00036; BZIP_BASIC; 1.
KW Transcription regulation; DNA-binding; Activator; Nuclear protein.
FT DNA_BIND 38 67 BASIC MOTIF.
FT DOMAIN 79 107 LEUCINE-ZIPPER.
SQ SEQUENCE 583 AA; 63885 MW; A4A96812D7F5C983 CRC64;

Query Match 54.4%; Score 43; DB 1; Length 583;
Best Local Similarity 58.8%; Pred. No. 34;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 SSPSSTKSSPSNVKAS 17
Db 214 SSISNSSSPSNVNGLS 230

RESULT 11
BRC2_HALRO STANDARD; PRT; 681 AA.
ID BRC2_HALRO
AC C01409;
DT 01-NOV-1997 (Rel. 35, Created)
```

DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE T-BOX CONTAINING PROTEIN 2 (AS-T2).  
 GN T2.  
 OS Halocynthia roretzi (Sea squirt).  
 OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;  
 OC Stolidobranchia; Pyuridae; Halocynthia.  
 OX NCBI\_TaxID=7729;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Gastrula;  
 RX MEDLINE=97115721; PubMed=8954744;  
 RA Yasuo H., Kobayashi M., Shimauchi Y., Satoh N.;  
 RT "The ascidian genome contains another T-domain gene that is expressed  
 RT in differentiating muscle and the tip of the tail of the embryo.";  
 RL Dev. Biol. 180:773-779(1996).  
 CC -1- FUNCTION: INVOLVED IN THE TRANSCRIPTIONAL REGULATION OF GENES  
 CC REQUIRED FOR MUSCLE DIFFERENTIATION. BINDS TO A PALINDROMIC  
 CC SITE (CALLED T SITE) AND ACTIVATES GENE TRANSCRIPTION WHEN BOUND  
 CC TO SUCH A SITE (BY SIMILARITY).  
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
 CC -1- TISSUE SPECIFICITY: DIFFERENTIATING MUSCLE AND TAILBUD TIP.  
 CC DEVELOPMENTAL STAGE: FIRST DETECTED IN THE VEGETAL HEMISPHERE AT  
 CC THE 32-CELL STAGE. EXPRESSION CONTINUES IN ENDOTHELIAL LINEAGE  
 CC UNTIL 110-CELL STAGE WHEN IT DECLINES HERE AND BECOMES EVIDENT  
 CC IN MUSCLE LINEAGE CELLS. EXPRESSION INCREASES THROUGHOUT  
 CC GASTRULATION AND NEURULATION IN DIFFERENTIATING MUSCLE CELLS  
 CC UNTIL LATE TAILBUD STAGE. EXPRESSION ALSO DETECTED IN THE TAILBUD  
 CC TIP FROM MID-GASTRULATION TO TAIL ELONGATION.  
 CC -1- SIMILARITY: CONTAINS A T-BOX DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D83265; BAA19689.1; -;  
 CC HSSP: P24781; IXHR.  
 CC InterPro: IPR001699; T-box.  
 CC Pfam: PF00907; T-box; 1.  
 CC PRINTS: PR00937; TBOX.  
 CC SMART: SM00425; TBOX; 1.  
 CC PROSITE: PS01283; TBOX\_1; 1.  
 CC PROSITE: PS01284; TBOX\_2; 1.  
 CC PROSITE: PS0252; TBOX\_3; 1.  
 CC Developmental protein; Transcription regulation; DNA-binding;  
 CC Nuclear protein; Activator.  
 CC DNA\_BIND 149 323 T-BOX.  
 CC DOMAIN 127 131 POLY-GLN.  
 FT SEQUENCE 681 AA; 76430 MW; 0DD8C8C0CD54A2E5 CRC64;  
 SO  
 Query Match 54.4%; Score 43; DB 1; Length 681;  
 Best Local Similarity 56.2%; Pred. No. 40;  
 Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
 Oy 1 SSPSSTKSSPSNVKSA 16  
 |||||:|:|:|:|:  
 Db 478 SSPSSDSNQSNVNAS 493  
 RESULT 12  
 ID VTA2\_XENLA STANDARD; PRT; 1807 AA.  
 AC P18709; Q91895;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE VITELLOGENIN A2 PRECURSOR (VTG A2) [CONTAINS: LIPOVITELLIN I;  
 DE LIPOVITELLIN II; PHOSVITIN; PHOSVETTE I; PHOSVETTE II].  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;  
 OC Xenopodidae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 16-28.  
 RX MEDLINE=87259958; PubMed=3601655;  
 RA Gerber-Huber S., Nardelli D., Haefliger J.-A., Cooper D.N., Givel F.,  
 RA Germond J.-E., Engel J., Green N.M., Wahli W.;  
 RT "Precursor-product relationship between vitellogenin and the yolk  
 RT proteins as derived from the complete sequence of a Xenopus  
 RT vitellogenin gene.";  
 RL Nucleic Acids Res. 15:4737-4760(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88058863; PubMed=3680202;  
 RA Nardelli D., Het Schip F.D., Gerber-Huber S., Haefliger J.A.,  
 RA Gruber M., Ab G., Wahli W.;  
 RT "Comparison of the organization and fine structure of a chicken and a  
 RT Xenopus laevis vitellogenin gene.";  
 RL J. Biol. Chem. 262:15377-15385(1987).  
 RN [3]  
 RP SEQUENCE OF 1291-1302.  
 RX MEDLINE=90278951; PubMed=2352275;  
 RA Wallace R.A., Hoch K.L., Carnevali O.;  
 RT "Placement of small lipovitellin subunits within the vitellogenin  
 RT precursor in Xenopus laevis.";  
 RL J. Mol. Biol. 213:407-409(1990).  
 RN [4]  
 RP SEQUENCE OF 1-72 FROM N.A.  
 RX MEDLINE=85062836; PubMed=6504705;  
 RA Walker P., Germond J.-E., Brown-Luedi M., Givel F., Wahli W.;  
 RT "Sequence homologues in the region preceding the transcription  
 RT initiation site of the liver estrogen-responsive vitellogenin and  
 RT apo-VLDLII genes.";  
 RL Nucleic Acids Res. 12:8611-8626(1984).  
 CC -1- FUNCTION: PRECURSOR OF THE MAJOR EGG-YOLK PROTEINS THAT ARE  
 CC SOURCES OF NUTRIENTS DURING EARLY DEVELOPMENT OF OVIPAROUS  
 CC ORGANISMS.  
 CC -1- TISSUE SPECIFICITY: PRODUCED BY THE LIVER, SECRETED INTO THE BLOOD  
 CC AND THEN SEQUESTERED BY RECEPTOR MEDIATED ENDOCYTOSIS INTO GROWING  
 CC OOCYTES, WHERE IT IS GENERALLY CLEAVED, GIVING RISE TO THE  
 CC RESPECTIVE YOLK COMPONENTS.  
 CC -1- INDUCTION: VITELLOGENIN IS SYNTHESIZED IN THE LIVER OF OVIPAROUS  
 CC VERTEBRATES IN RESPONSE TO STEROID (ESTROGEN) INDUCTION.  
 CC -1- MISCELLANEOUS: THE SERINE-RICH PORTION OF VITELLOGENIN ENCODES  
 CC TO A LEVEL OF ABOUT 80%.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Y00354; CAA68433.1; -;  
 CC EMBL: M18061; AAA49982.1; -;  
 CC EMBL: X00205; CAA25028.1; -;  
 CC PIR: S03124; S03124.  
 CC PIR: S10625; S10625.  
 CC InterPro: IPR001747; Vitellogenin\_N.  
 CC InterPro: IPR001846; Vwd.  
 CC Pfam: PF01347; Vitellogenin\_N; 1.  
 CC Pfam: PF00094; vwd; 1.  
 CC SMART: SM00216; VWD; 1.  
 CC Storage protein; Phosphorylation; Signal; Glycoprotein.  
 KW SIGNAL 1 15  
 FT CHAIN 16 1807 VITELLOGENIN A2.  
 FT CHAIN 16 ? LIPOVITELLIN I.  
 FT CHAIN 16 ?

```
FT CHAIN ? 1290 PHOSVITIN (OR PHOSVETTES I + II).
FT CHAIN 1291 ? LIPOVITELLIN II.
FT DOMAIN 1126 1321 SER-RICH.
FT DOMAIN 1126 1163 POLY-SER.
FT DOMAIN 1179 1184 POLY-SER.
FT DOMAIN 1205 1232 POLY-SER.
FT DOMAIN 1263 1275 POLY-SER.
FT DOMAIN 1309 1321 POLY-SER.
FT CARBOHYD 1094 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 662 662 V -> I.
FT VARIANT 958 958 A -> T.
FT VARIANT 1572 1572 E -> K.
FT CONFLICT 172 72 I -> V (IN REF. 4).
SQ SEQUENCE 1807 AA; 201544 MW; D109B9F568147742 CRC64;

Query Match 54.4%; Score 43; DB 1; Length 1807;
Best Local Similarity 58.8%; Pred. No. 1.le+02;
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 SSPSTKSSPSNVKAS 17
DB 1145 SSSSSSSSSSSSSSS 1161

RESULT 13
IE18_PRIVIF STANDARD; PRT; 1461 AA.
AC P11675;
RA de Robertis E.M.;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE IMMEDIATE-EARLY PROTEIN IE180.
GN IE.
OS Pseudorabies virus (strain Indiana-Funkhauser / Becker) (PRV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=31523;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89315207; PubMed=2546124;
RA "Cheung A.K.;
RT "DNA nucleotide sequence analysis of the immediate-early gene of
pseudorabies virus.";
RL Nucleic Acids Res. 17:4637-4646(1989).
RN [2]
RP REVISIONS.
RA Cheung A.K.;
RL Submitted (NOV-1989) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE
OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING
OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS.
CC -!- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.
CC -!- PTM: A LONG STRETCH OF SERINE RESIDUES MAY BE A MAJOR SITE OF
PHOSPHORYLATION.
CC -!- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: X15120; CAA33214.1; -.
DR PIR; S04713; EDBEIF.
KW Early protein; Transcription regulation; Trans-acting factor;
KW DNA-binding; Phosphorylation; Nuclear protein.
FT DOMAIN 390 405 POLY-SER.
FT DOMAIN 958 966 POLY-SER.
SQ SEQUENCE 1461 AA; 149833 MW; 7F31E7ABEA03B208 CRC64;
```

```
Query Match 53.8%; Score 42.5; DB 1; Length 1461;
Best Local Similarity 64.7%; Pred. No. 1.le+02;
Matches 11; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

QY 1 SSPSTKSSPSNVKAS 17
DB 853 SSPASTKSG-SSTKSS 868

RESULT 14
RRA_XENLA STANDARD; PRT; 458 AA.
ID RRA_XENLA
AC P51126;
DT 01-OCT-1996 (Rel. 34, Created)
DT 20-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RETINOIC ACID RECEPTOR ALPHA (RAR-ALPHA).
GN RARA OR NR1B1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92196110; PubMed=1312717;
RA Blumberg B., Mangelsdorf D.J., Dyck J.A., Bittner D.A., Evans R.M.,
de Robertis E.M.;
RT "Multiple retinoid-responsive receptors in a single cell: families of
retinoid 'x' receptors and retinoic acid receptors in the Xenopus
egg.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:2321-2325(1992).
CC -!- FUNCTION: THIS IS A RECEPTOR FOR RETINOIC ACID. THIS METABOLITE
HAS PROFOUND EFFECTS ON VERTEBRATE DEVELOPMENT. RETINOIC ACID IS
A MORPHOGEN AND IS A POWERFUL TERATOGEN. THIS RECEPTOR CONTROLS
CELLS FUNCTIONS BY DIRECTLY REGULATING GENE EXPRESSION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DEVELOPMENTAL STAGE: IT IS SYNTHESIZED DURING OOGENESIS AND
PERSISTS IN THE CLEAVING EMBRYO AT APPROXIMATELY CONSTANT LEVELS
UNTIL IT IS DEGRADED JUST BEFORE GASTRULATION.
CC -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
NR1 SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: L11445; -. NOT_ANNOTATED_CDS.
DR HSSP; P10826; 1HRA.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Steroidhormone_rcptor.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00047; STROIDFINGER.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR01292; RETNOICACIDR.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger; Multigene family.
FT DOMAIN 1 87 MODULATING (BY SIMILARITY).
FT NUCLEO-BIND 88 153 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 88 108 C4-TYPE.
FT ZN_FING 124 148 C4-TYPE.
```

Job time: 148 sec

FT DOMAIN 154 199 HINGE.  
FT DOMAIN 200 418 LIGAND-BINDING (BY SIMILARITY).  
SQ SEQUENCE 458 AA; 50573 MW; 4D80BB18678B1E17 CRC64;

Query Match 53.2%; Score 42; DB 1; Length 458;  
Best Local Similarity 90.0%; Pred. No. 37;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 SPSSTKSSPS 11  
      |||||  
Db 445 SPSSTHSSPS 454

RESULT 15  
ID YN23\_YEAST STANDARD; PRT; 503 AA.  
AC P53832;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE HYPOTHETICAL 52.3 KDA PROTEIN IN MRPL10-ERG24 INTERGENIC REGION  
DE PRECURSOR.  
GN YNL283C OR N0583.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Messenguy F., Dubois E., Vlerendeels F., Scherens B., Pierard A.,  
RA Glansdorff N.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: TO S.POMBE SPBC3D5.14C.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; Z71559; CAA96195.1; -  
CC SGD; S0005227; YNL283C.  
CC InterPro: IPR002889; WSC.  
CC Pfam; PF01822; WSC; 1.  
CC SMART; SM00321; WSC; 1.  
CC Hypothetical protein; Transmembrane; Signal.  
FT SIGNAL 1 23  
FT CHAIN 1 23  
FT TRANSMEM 61 81  
FT TRANSMEM 326 346  
FT CARBOHYD 394 394  
FT CARBOHYD 429 429  
FT CARBOHYD 475 475  
FT CARBOHYD 483 483  
FT CARBOHYD 498 498  
FT CARBOHYD 499 499  
SQ SEQUENCE 503 AA; 52292 MW; F2392A73C5CBAB50 CRC64;

Query Match 53.2%; Score 42; DB 1; Length 503;  
Best Local Similarity 52.9%; Pred. No. 41;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SSPSSTKSSPSNVKAS 17  
      ||:|||||  
Db 203 SSTTSTSSPSSTTSST 219

Search completed: January 29, 2002, 11:16:19

us-09-763-397a-22.rsp

Mon Feb 4 15:23:44 2002

\_\_\_\_\_



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 11:12:20 ; Search time 62.77 Seconds  
(without alignments)  
39.615 Million cell updates/sec

Title: US-09-763-397A-22

Perfect score: 79  
Sequence: 1 SSPSSTKSSPSNVKSAS 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL\_17.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_invertebrate.\*  
14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	68.5	86.7	171	5 Q26002	Q26002 plasmodium
2	68.5	86.7	171	5 Q26003	Q26003 plasmodium
3	68.5	86.7	782	5 Q25730	Q25730 plasmodium
4	68.5	86.7	782	5 Q25675	Q25675 plasmodium
5	68.5	86.7	782	5 Q26007	Q26007 plasmodium
6	68.5	86.7	782	5 Q9U431	Q9U431 plasmodium
7	68.5	86.7	782	5 Q9U430	Q9U430 plasmodium
8	68.5	86.7	782	5 Q9U429	Q9U429 plasmodium
9	68.5	86.7	782	5 Q9U414	Q9U414 plasmodium
10	65.5	82.9	782	5 Q26104	Q26104 plasmodium
11	62.5	79.1	292	5 Q25762	Q25762 plasmodium
12	50	63.3	2559	5 Q9V7X3	Q9V7X3 drosophila
13	49	62.0	493	11 Q9JKN6	Q9JKN6 mus musculus
14	48	60.8	328	10 O81138	O81138 arabidopsis
15	47	59.5	98	5 Q9NPH0	Q9NPH0 plasmodium
16	47	59.5	215	13 Q98TQ9	Q98TQ9 gallus gall
17	47	59.5	497	2 Q9PAB6	Q9PAB6 xylella fas
18	47	59.5	795	5 Q9YIJ5	Q9YIJ5 dictyosteli
19	46	58.2	243	11 Q9QY63	Q9QY63 mus musculus

20	46	58.2	244	4 Q9Y4S1	Q9Y4S1 homo sapien
21	46	58.2	312	10 Q9LHF0	Q9LHF0 arabidopsis
22	46	58.2	1514	4 Q9P2F8	Q9P2F8 homo sapien
23	45	57.0	211	10 Q9LPE8	Q9LPE8 arabidopsis
24	45	57.0	349	3 Q74947	Q74947 schizosacch
25	45	57.0	383	5 Q9GWA1	Q9GWA1 leishmania
26	45	57.0	590	11 O88337	O88337 rattus norv
27	45	57.0	625	11 Q9ROU2	Q9ROU2 mus musculu
28	45	57.0	626	11 O88295	O88295 mus musculu
29	45	57.0	626	11 O88529	O88529 mesocricetu
30	45	57.0	626	11 O88810	O88810 rattus norv
31	45	57.0	632	11 Q9WTL8	Q9WTL8 mus musculu
32	45	57.0	633	13 Q918T7	Q918T7 gallus gall
33	45	57.0	766	12 Q9DMF7	Q9DMF7 rat cytomeg
34	45	57.0	2406	5 Q9VXM5	Q9VXM5 drosophila
35	45	57.0	2559	5 O44113	O44113 drosophila
36	45	57.0	2559	5 O44381	O44381 drosophila
37	44	55.7	103	4 O60372	O60372 homo sapien
38	44	55.7	163	4 Q9NPN4	Q9NPN4 homo sapien
39	44	55.7	214	10 O49780	O49780 arabidopsis
40	44	55.7	229	10 O81864	O81864 arabidopsis
41	44	55.7	254	4 Q9UIX6	Q9UIX6 homo sapien
42	44	55.7	267	10 Q9M9U2	Q9M9U2 arabidopsis
43	44	55.7	311	4 Q9NWX1	Q9NWX1 homo sapien
44	44	55.7	337	5 O16139	O16139 lytechinus
45	44	55.7	398	3 Q9HED1	Q9HED1 neurospora

ALIGNMENTS

RESULT 1					
Q26002	AC	Q26002;	PRELIMINARY;	PRT;	171 AA.
DT	01-NOV-1996	(TREMBLrel. 01, Created)			
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)			
DT	01-NOV-1998	(TREMBLrel. 08, Last annotation update)			
DE	RHOPTRY PROTEIN (FRAGMENT).				
GN	RAP-1.				
OS	Plasmodium falciparum.				
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
OX	NCBI_TaxID=5833;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=SIERRA LEONE;				
RX	MEDLINE=92244303; PubMed=1574089;				
RA	Howard R.F.;				
RT	"The sequence of the p82 rhopty protein is highly conserved between				
RL	two Plasmodium falciparum isolates";				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=SIERRA LEONE;				
RX	MEDLINE=93293317; PubMed=7685740;				
RA	Howard R.F., Jensen J.B., Franklin H.L.;				
RT	"Reactivity profile of human anti-82-kilodalton rhopty protein				
RL	antibodies generated during natural infection with Plasmodium				
RN	falciparum";				
DR	Infect. Immun. 61:2960-2965(1993).				
RL	EMBL: L10322; AAB59202.1; -.				
FT	NON_TER 1				
FT	NON_TER 171				
SQ	SEQUENCE 171 AA; 18510 MW; 8BF5709D8D3DB844 CRC64;				

Query Match	86.7%;	Score 68.5;	DB 5;	Length 171;
Best Local Similarity	94.4%;	Pred. No. 0.0015;		
Matches 17;	Conservative 0;	Mismatches 0;	Indels 1;	Gaps 1;
QY	1 SSPSSTK-SSPSNVKSAS 17			
Db	22 SSPSSTKSSPSNVKSAS 39			

```

RESULT 2
Q26003 ID Q26003 PRELIMINARY; PRT; 171 AA.
AC Q26003;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE RHOPTRY PROTEIN (FRAGMENT).
GN RAP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TANZANIA I/CDC;
MEDLINE=92244303; PubMed=1574089;
RA Howard R.F.;
RT "The sequence of the p82 rhoptry protein is highly conserved between two Plasmodium falciparum isolates."
RL Mol. Biochem. Parasitol. 51:327-330(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-TANZANIA I/CDC;
MEDLINE=93293317; PubMed=7685740;
RA Howard R.F.; Jensen J.B.; Franklin H.L.;
RT "Reactivity profile of human anti-82-kilodalton rhoptry protein antibodies generated during natural infection with Plasmodium falciparum."
RT Infect. Immun. 61:2960-2965(1993).
RL EMBL; L10323; AAA29742.1; -.
DR EMBL; M80807; AAA29717.1; -.
FT NON_TER 1
FT NON_TER 171
SQ SEQUENCE 171 AA; 18531 MW; 192CBACFC68970CB8 CRC64;

Query Match 86.7%; Score 68.5; DB 5; Length 171;
Best Local Similarity 94.4%; Pred. No. 0.0015; Mismatches 0; Indels 1; Gaps 1;
Matches 17; Conservative 0;

QY 1 SSPSTK-SSPSNVKSAS 17
||||| |||||||
DB 22 SSPSTKSSPSNVKSAS 39

RESULT 3
Q25730 ID Q25730 PRELIMINARY; PRT; 782 AA.
AC Q25730;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE RHOPTRY ASSOCIATED PROTEIN-1.
GN RAP-1.
OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FC27;
RA Shi Y.; Lai A.A.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20985; AAA63681.1; -.
SQ SEQUENCE 782 AA; 90023 MW; F69E26A2A564C8EA CRC64;

Query Match 86.7%; Score 68.5; DB 5; Length 782;
Best Local Similarity 94.4%; Pred. No. 0.0059; Mismatches 0; Indels 1; Gaps 1;
Matches 17; Conservative 0;

QY 1 SSPSTK-SSPSNVKSAS 17
||||| |||||||

```

```

Db 145 SSPSTKSSPSNVKSAS 162

RESULT 4
Q25875 ID Q25875 PRELIMINARY; PRT; 782 AA.
AC Q25875;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE PR86 RHOPTRY PRECURSOR PROTEIN.
GN PR86.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HONDURAS I CDC;
MEDLINE=90348711; PubMed=2200961;
RA Ridley R.G.; Takacs B.; Lahm H.W.; Delves C.J.; Goman M.; Certa V.;
RA Matile H.; Woollett G.R.; Scaife J.G.;
RT "Characterisation and sequence of a protective rhoptry antigen from Plasmodium falciparum."
RT Mol. Biochem. Parasitol. 41:125-134(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-HONDURAS I CDC;
MEDLINE=92244303; PubMed=1574089;
RA Howard R.F.;
RT "The sequence of the p82 rhoptry protein is highly conserved between two Plasmodium falciparum isolates."
RL Mol. Biochem. Parasitol. 51:327-330(1992).
DR EMBL; M80807; AAA29717.1; -.
SQ SEQUENCE 782 AA; 90096 MW; D1AD099862528D42 CRC64;

Query Match 86.7%; Score 68.5; DB 5; Length 782;
Best Local Similarity 94.4%; Pred. No. 0.0059; Mismatches 0; Indels 1; Gaps 1;
Matches 17; Conservative 0;

QY 1 SSPSTK-SSPSNVKSAS 17
||||| |||||||
DB 145 SSPSTKSSPSNVKSAS 162

RESULT 5
Q26007 ID Q26007 PRELIMINARY; PRT; 782 AA.
AC Q26007;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE RHOPTRY ASSOCIATED PROTEIN 1 (ROPI) PRECURSOR (FRAGMENT).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=90348711; PubMed=2200961;
RA Ridley R.G.; Takacs B.; Lahm H.W.; Delves C.J.; Goman M.; Certa V.;
RA Matile H.; Woollett G.R.; Scaife J.G.;
RT "Characterisation and sequence of a protective rhoptry antigen from Plasmodium falciparum."
RL Mol. Biochem. Parasitol. 41:125-134(1990).
DR EMBL; M32853; AAA29753.1; -.
KW SIGNAL.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 >782 POTENTIAL.
FT NON_TER 782
SQ SEQUENCE 782 AA; 90080 MW; 2F2E240D40A4C902 CRC64;

Query Match 86.7%; Score 68.5; DB 5; Length 782;

```

1 SSPSSTK-SSPSSNVKSSAS 17

97	I		SSPSSTKSSSPSNVKSAS	162
			SSPSSTKSSSPSNVKSAS	162
Db	145	SSPSSTKSSSPSNVKSAS		162

SA	Saul A.;
RT	"Efficacy of vaccines containing Rhoptyr-Associated Proteins RAP1 and
RA	RAP2 of <i>Plasmodium falciparum</i> in Saimiri boliviensis monkeys. ";
RL	Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
DR	EMBL; AF205282; AAF23403.1; -
RL	NON_TER 782 782
SE	SEQUENCE 782 AA: 90066 MW: B344948D5806F7DC CRC64:
SQ	

Mismatches	0; Mismatches	0; Indels	1; Gaps
Matches	17; Conservative	0; Indels	1; Gaps

Db 145 SSPSSTKSSSPSNVKSAS 162

OX	NCBI_TaxID=5833;	
RN	[1]	
RN	SEQUENCE FROM N.A.	
RC	STRAIN=EUP;	
RC	Saul A.;	
RT	Efficacy of vaccines containing Rhoptyr-Associated Proteins	
RT	RAP2 of <i>Plasmodium falciparum</i> in <i>Salimiri boliviensis</i> monkeys	
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBD databases.	
RL	EMBL: AF205283; AAF23404.1;	
RD	NON_TER	782
RD	SEQUENCE	782 AA;
RD	782	90131 MW; A0FCB64529C34E59 CRC64;
RD	SEQUENCE	782 AA;
RD	782	90131 MW; A0FCB64529C34E59 CRC64;

<b>Matches</b>	<b>17;</b>	<b>Conservative</b>	<b>0;</b>	<b>Mismatches</b>	<b>0;</b>	<b>Indels</b>	<b>1;</b>	<b>Gaps</b>	<b>1;</b>
----------------	------------	---------------------	-----------	-------------------	-----------	---------------	-----------	-------------	-----------

Db 145 SSPSTKSSSPSNVKSAS 162

Qy	1	SSPSTK-SSPSNVKSAS	17
Dd	145	SSPSTKSSSPSNVKSAS	162

Q9U414 ID Q9U414 PRELIMINARY: PRT: 782 AA.

Q9UQ14	PRELIMINARY; PRT; 782 AA.
AC	Q9UQ14;
DT	01-MAY-2000 (TrEMBLrel. 13, Created)
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT	01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE	RHOPIRY ASSOCIATED PROTEIN-1.
GN	RAP-1.
OS	Plasmodium falciparum.
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium
OX	NCBI_TaxID=5833;
RN	[1]
RP	SEQUENCE FROM N. A.
RC	STRAIN=FC1/HN;
RA	Li X. R., Luo S. H., Yu X. B., Shan Z. X., Ma C. L.;
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RQ	EMBL; AF206631; AAF15365.1; -
SR	SEQUENCE 782 AA; 90082 MW; 8E1F4CF2883903FD CRC64;

Query Match 86.7%; Score 68.5; DB 5; Length 782;  
Best Local Similarity 94.4%; Pred. No. 0.0059;  
Matches 17; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 SSPSSTK-SSPSNVKSAS 17  
 ||||| |||||  
 Db 145 SSPSSTKSSSPSNVKSAS 162

RESULT 10  
Q26104  
ID Q26104 PRELIMINARY; PRT; 782 AA.

Q26907	01-NOV-1996	(TrEMBLrel. 01, Created)
DT	01-NOV-1996	(TrEMBLrel. 01, Last sequence update)
DT	01-JUN-2000	(TrEMBLrel. 14, Last annotation update)
DE		RHOPTRY ASSOCIATED PROTEIN 1.

DE KHOPRI ASSOCIATED PROTEIN 1.  
GN RAP-1.  
OS Plasmodium reichenowi.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

Mon Feb 4 15:23:44 2002

us-09-763-397a-22.rspt

Ephydroidea; Drosophilidae; Drosophila.

Page 4

OX NCBI\_TaxID=5854;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Shi Y., Lai A.A.;  
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U20986; AAA63682.1; -  
SQ SEQUENCE 782 AA; 90160 MW; DE6D1BE2FAC308A9 CRC64;

Query Match 82.9%; Score 65.5; DB 5; Length 782;  
Best Local Similarity 88.9%; Pred. No. 0.018; Indels 1; Gaps 1;  
Matches 16; Conservative 1; Mismatches 0;

OY 1 SSPSSTK-SSPSNVKAS 17  
Db 145 SSPSSTKSSPSNVKTAS 162  
||||||| ||||||| ||

RESULT 11  
Q25762 PRELIMINARY; PRT; 292 AA.  
AC Q25762;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
DE STRAIN INDD RHOPTRY-ASSOCIATED PROTEIN 1 (RAP-1) (FRAGMENT).  
GN RAP-1.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-INDD (INDIAN ISOLATE D);  
RX MEDLINE=92244303; PubMed=1574089;  
RA Howard R.F.;  
RT "The sequence of the p82 rhoptry protein is highly conserved between two Plasmodium falciparum isolates."  
RL Mol. Biochem. Parasitol. 51:327-330(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-INDD (INDIAN ISOLATE D);  
RX MEDLINE=96379224; PubMed=8784775;  
RA Howard R.F., Peterson C.;  
RT "Limited RAP-1 sequence diversity in field isolates of Plasmodium falciparum."  
RL Mol. Biochem. Parasitol. 77:95-98(1996).  
DR EMBL; U41074; AAC47090.1; -  
FT NON\_TER 292 1  
FT NON\_TER 292 1  
SQ SEQUENCE 292 AA; 32345 MW; A35C979C3FEC3019 CRC64;

Query Match 79.1%; Score 62.5; DB 5; Length 292;  
Best Local Similarity 88.9%; Pred. No. 0.022; Indels 1; Gaps 1;  
Matches 16; Conservative 0; Mismatches 1;

OY 1 SSPSSTK-SSPSNVKAS 17  
Db 139 SSPSSTKSSPSNVKAS 156  
||||||| ||||||| ||

RESULT 12  
Q9V7X3 PRELIMINARY; PRT; 2559 AA.  
AC Q9V7X3;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE RHOGEF2 PROTEIN.  
GN RHOGEF2 OR CG9635.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA April J.F., Adayani A., An H.-J., Andrews-pfannkuch C., Baldwin D.,  
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,  
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Moberly C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J., Yao Q.A.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Zhang G., Zheng L.,  
RA Williams J., Yeh R.-F., Zaveri J.S., Zhan M., Zhou X., Zhu X., Smith H.O.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zou Y., Zou Y.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster."  
RL Science 287:2185-2195(2000).  
DR EMBL; AE003804; AAF57915.1; -  
DR HSP; Q12923; 3PDZ.  
DR FlyBase; FBgn0023172; RhoGEF2.  
DR InterPro; IPR001525; C5\_DNA\_meth.  
DR InterPro; IPR002219; DAG\_PE-bind.  
DR InterPro; IPR001478; PDZ.  
DR InterPro; IPR001849; PH.  
DR InterPro; IPR000219; RhoGEF.  
DR Pfam; PF00130; DAG\_PE-bind; 1.  
DR Pfam; PF00595; PDZ; 1.  
DR Pfam; PF00169; PH; 1.  
DR Pfam; PF00621; RhoGEF; 1.  
DR SMART; SM00109; C1; 1.  
DR SMART; SM00228; PDZ; 1.  
DR SMART; SM00233; PH; 1.  
DR SMART; SM00325; RhoGEF; 1.  
DR PROSITE; PS00094; C5\_MTASE\_1; UNKNOWN\_1.  
DR PROSITE; PS00479; DAG\_PE\_BIND\_DOM\_1; 1.  
DR PROSITE; PS00081; DAG\_PE\_BIND\_DOM\_2; 1.  
DR PROSITE; PS0106; PDZ; 1.  
DR SEQUENCE 2559 AA; 280805 MW; 07725DE04982605C CRC64;

Query Match 63.3%; Score 50; DB 5; Length 2559;  
Best Local Similarity 76.9%; Pred. No. 14;  
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 PSSTKSSPSNVKS 15  
||||| :||||| ||

Db 645 PSSTDNSPSNAKS 657

RESULT 13

Q9JKN6 PRELIMINARY; PRT; 493 AA.

AC Q9JKN6; (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-OCT-2000 (TREMBlrel. 17, Last annotation update)

DE VENTRAL NEURON-SPECIFIC PROTEIN 1 NOVAL (FRAGMENT).

GN NOVAL.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CBA/J; TISSUE=BRAIN;

RA Ward-Bailey P.F., Johnson K.R.;

RT "Neuromuscular ataxia: a new spontaneous mutation in the mouse."

RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF232828; AAF35907.1;

DR MGD; MGI:104297; NOVAL.

DR InterPro; IPR000958; KH.

DR Pfam; PF00013; KH-domain; 3.

DR SMART; SM00322; KH; 3.

FT NON\_TER 1

SQ SEQUENCE 493 AA; 50302 MW; 7D09E8A55B0A7817 CRC64;

Query Match 62.0%; Score 49; DB 11; Length 493;

Best Local Similarity 58.8%; Pred. No. 4.8;

Matches 10; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 SSPSTKSSPSNVKSAS 17

|||||:|||||:

Db 139 SSPTTKSSPSDPMTTS 155

RESULT 14

O81138 PRELIMINARY; PRT; 328 AA.

AC O81138;

DT 01-NOV-1998 (TREMBlrel. 08, Created)

DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE AP2 DOMAIN FAMILY TRANSCRIPTION FACTOR HOMOLOG.

GN ABI4 OR AT2G40220.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RX MEDLINE=9829844; Pubmed=9634591;

RA Finkelstein R.R., Wang M.L., Lynch T.J., Rao S., Goodman H.M.;

RT "The Arabidopsis abscisic acid response locus ABI4 encodes an APETALA

2 domain protein."

RL Plant Cell 10:1043-1054(1998).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=99225673; Pubmed=10207155;

RA Wang M.L., Belmonte S., Kim U., Dolan M., Morris J.W., Goodman H.M.;

RT "A cluster of ABA-regulated genes on Arabidopsis thaliana BAC

T07M07."

RL Genome Res. 9:325-333(1999).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RX MEDLINE=20083487; Pubmed=10617197;

RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,  
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,  
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,  
RA Salzberg S.L., Fraser C.M., Venter J.C.;  
RT "Sequence and analysis of chromosome II of Arabidopsis thaliana.";  
RL Nature 402:761-768(1999).  
DR EMBL: AF040959; AAC39489.1;  
DR EMBL: AF085279; AAD25937.1;  
DR EMBL: AC018721; AAF18736.1;  
DR HSP; O80337; 2GCC.  
DR TRANSFAC; T02642;  
DR InterPro; IPR001471; AP2-domain.  
DR Pfam; PF00847; AP2-domain; 1.  
DR PRINTS; PR00367; ETHRSPELEMT.  
DR ProDom; PD001423; AP2-domain; 1.  
DR SMART; SM00380; AP2; 1.  
SQ SEQUENCE 328 AA; 35671 MW; B6B710475E34D08F CRC64;

Query Match 60.8%; Score 48; DB 10; Length 328;

Best Local Similarity 64.7%; Pred. No. 4.8;

Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 SSPSTKSSPSNVKSAS 17

|||||:|||||:

Db 113 SSPSSVSSSSSVSAAS 129

RESULT 15

O9NFH0 PRELIMINARY; PRT; 98 AA.

AC O9NFH0;

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE HYPOTHETICAL 9.6 KDA PROTEIN (FRAGMENT).

OS Plasmodium falciparum (isolate 3D7).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI\_TaxID=36329;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=3D7;

RA Spielmann T., Beck H.;

RT "Analysis of stage specific transcription in Plasmodium falciparum

RT reveals a set of genes exclusively transcribed in ring stage

RT parasites."

RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AJ290927; CAB92936.1;

KW Hypothetical protein.

FT NON\_TER 1

FT NON\_TER 98

SQ SEQUENCE 98 AA; 9590 MW; BC1857EB0D303042 CRC64;

Query Match 59.5%; Score 47; DB 5; Length 98;

Best Local Similarity 62.5%; Pred. No. 2.4;

Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 2 SSPSTKSSPSNVKSAS 17

|||||:|||||:

Db 14 SSPSSSSSPSSSSSS 29

Search completed: January 29, 2002, 11:15:52  
Job time: 212 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 10:59:18 ; Search time 66.28 Seconds  
(without alignments)  
18.999 Million cell updates/sec

Title: US-09-763-397A-23

Perfect score: 84

Sequence: 1 LATRLMKFKAEIRDFF 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_1101.\*  
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	84	100.0	17	AAV70299	Plasmodium falciparum
2	84	100.0	350	AAV70278	Recombinant vaccinia
3	84	100.0	782	AAV06991	Polypeptide antigen
4	45	53.6	190	AAV82568	S. epidermidis ope
5	45	53.6	1572	AAV82906	S. epidermidis ope
6	43	51.2	86	AAV4822	Gene 36 human secr
7	43	51.2	86	AAV4822	Human secreted pro
8	43	51.2	98	AAV57133	Human prostate can
9	43	51.2	297	AAV31171	Arabidopsis thalia
10	43	51.2	335	AAV31170	Arabidopsis thalia
11	43	51.2	353	AAV31169	Arabidopsis thalia

12	42	50.0	186	21	AAV04103	Arabidopsis thalia
13	42	50.0	190	21	AAV04102	Arabidopsis thalia
14	42	50.0	203	21	AAV04101	Arabidopsis thalia
15	41	48.8	300	21	AAV07741	Arabidopsis thalia
16	41	48.8	300	21	AAV38362	Arabidopsis thalia
17	41	48.8	302	21	AAV07740	Arabidopsis thalia
18	41	48.8	302	21	AAV38361	Arabidopsis thalia
19	41	48.8	319	21	AAV07739	Arabidopsis thalia
20	41	48.8	319	21	AAV38360	Arabidopsis thalia
21	41	48.8	2325	17	AAV05590	Maize acetyl CoA c
22	41	48.8	2325	19	AAV56736	Maize ACCase enzym
23	41	48.8	2325	21	AAV44687	Maize acetyl CoA c
24	40	47.6	333	21	AAV05123	Arabidopsis thalia
25	40	47.6	333	21	AAV50040	Arabidopsis thalia
26	40	47.6	338	21	AAV05122	Arabidopsis thalia
27	40	47.6	338	21	AAV50039	Arabidopsis thalia
28	40	47.6	356	21	AAV05121	Arabidopsis thalia
29	40	47.6	356	21	AAV50038	Arabidopsis thalia
30	40	47.6	534	21	AAV77232	Maize cytochrome p
31	40	47.6	534	21	AAV44726	Maize cytochrome p
32	39	46.4	247	20	AAV06357	Chaetomium brasill
33	39	46.4	247	21	AAV14870	Chaetomium brasill
34	39	46.4	247	21	AAV84335	Amino acid sequenc
35	39	46.4	375	21	AAV23212	Arabidopsis thalia
36	39	46.4	423	21	AAV23211	Arabidopsis thalia
37	39	46.4	485	21	AAV23210	Arabidopsis thalia
38	39	46.4	1147	14	AAV41199	CAI antigen, Heli
39	39	46.4	1338	19	AAV80600	Helicobacter pylor
40	38.5	45.8	133	21	AAV33094	Pinus radiata tran
41	38	45.2	281	19	AAV72190	HSV-2 strain SB5 C
42	38	45.2	354	19	AAV72122	HSV-2 strain SB5 C
43	38	45.2	354	19	AAV72045	HSV-2 strain SB5 C
44	38	45.2	475	21	AAV50384	Arabidopsis thalia
45	38	45.2	539	21	AAV50383	Arabidopsis thalia

#### ALIGNMENTS

RESULT 1

AAV70299

ID AAV70299 standard; peptide; 17 AA.

XX

AC AAV70299;

XX

XX

DT 06-JUN-2000 (first entry)

XX

DE Plasmodium falciparum RAP-1 antigenic epitope, P605.

XX

Recombinant protein; CDC/NIAIDVAC-1; multivalent; malaria; vaccine;

T-cell epitope; tetanus toxoid; antigenic epitope; treatment;

circumsporozoite protein; CSP; sporozoite surface protein-2; SSP-2;

liver stage antigen-1; LSA-1; merozoite surface protein-1; MSP-1; MSP-2;

apical membrane antigen-1; AMA-1; erythrocyte binding antigen-175;

EBA-175; rhoptry associated protein-1; RAP-1; gamete specific antigen;

Pf27; antiparasitic; prevention; anti-CDC/NIAIDVAC-1 antibody.

XX

OS Plasmodium falciparum.

XX

PN WO200011179-A1.

XX

PD 02-MAR-2000.

XX

PF 19-AUG-1999; 99WO-US18869.

XX

PR 21-AUG-1998; 98US-0097703.

XX

PA (NAIM-) NAT INST IMMUNOLOGY.

XX

PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX

PI Lal AA, Shi YP, Hasnain SE;

XX

XX WPI; 2000-237654/20.

DR

us-09-763-397a-23.rag

Mon Feb 4 15:23:44 2002

```

XX PA (NAIM-) NAT INST IMMUNOLOGY.
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX PI Lal AA, Shi YP, Hasnain SE;
XX DR WPI: 2000-237654/20.
XX DR N-PSDB; RAZ51336.
XX
XX Novel recombinant protein as vaccine for treating malarial infection
XX PT comprises antigenic peptides obtained from different stages of
XX PT plasmodium falciparum life cycle -
XX
XX Claim 2: Page 17; 52pp; English.
XX
XX The present sequence is the antigenic epitope P605, derived from
XX CC rhoptry associated protein-1 (RAP-1) of the asexual blood stage of
XX CC Plasmodium falciparum. It is used in the construction of recombinant
XX CC protein CDC/NIIMALVAC-1, which is a multivalent, multistage malarial
XX CC vaccine. The recombinant protein comprises, melittin signal peptide,
XX CC (His)6 tag, T-cell epitope from tetanus toxoid and 21 antigenic epitopes
XX CC from circumsporozoite protein (CSP), sporozoite surface protein-2
XX CC (SSP-2), liver stage antigen-1 (LSA-1), merozoite surface protein-1
XX CC (MSP-1), MSP-2, apical membrane antigen-1 (AMA-1), erythrocyte binding
XX CC antigen-175 (EBA-175), rhoptry associated protein-1 (RAP-1) and gamete
XX CC specific antigen, Pf27. These epitopes were obtained at different stages
XX CC of the life cycle of P. falciparum. CDC/NIIMALVAC-1 vaccine has
XX CC antiparasitic activity and can be used for treatment and prevention of
XX CC malarial infections. Anti-CDC/NIIMALVAC-1 antibodies can be used for
XX CC detecting P. falciparum in biological samples.
XX
XX Sequence 17 AA:
XX
XX Query Match 100.0%; Score 84; DB 21; Length 17;
XX Best Local Similarity 100.0%; Pred. No. 1.4e-07;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 LATRLMKKFKAEIRDFF 17
XX Db 1 latrlmkkfkaeirdff 17
XX
XX RESULT 2
XX ID AAY70278 standard; Protein: 350 AA.
XX AC AAY70278;
XX
XX DT 06-JUN-2000 (first entry)
XX DE Recombinant vaccine CDC/NIIMALVAC-1.
XX
XX KW Recombinant protein; CDC/NIIMALVAC-1; multivalent; malaria; vaccine;
XX KW T-cell epitope; tetanus toxoid; antigenic epitope; treatment;
XX KW circumsporozoite protein; CSP; sporozoite surface protein-2; SSP-2;
XX KW liver stage antigen-1; LSA-1; merozoite surface protein-1; MSP-1; MSP-2;
XX KW apical membrane antigen-1; AMA-1; erythrocyte binding antigen-175;
XX KW EBA-175; rhoptry associated protein-1; RAP-1; gamete specific antigen;
XX KW Pf27; antiparasitic; prevention; anti-CDC/NIIMALVAC-1 antibody;
XX KW honey bee.
XX
XX OS Chimeric - Apis sp.
XX OS Chimeric - Clostridium tetani.
XX OS Chimeric - Plasmodium falciparum.
XX
XX FH Key Location/Qualifiers
XX FT Peptide 1..22
XX FT /label= Melittin signal peptide
XX FT /note= "Derived from Honey bee"
XX FT Protein 23..350
XX FT /label= Mature_CDC/NIIMALVAC-1
XX FT /note= "Recombinant multivalent malarial vaccine"
XX
XX WO200011179-A1.
XX
XX PD 02-MAR-2000.
XX
XX PF 19-AUG-1999; 99WO-US18869.
XX
XX PR 21-AUG-1998; 98US-0097703.

```

```

XX PA (NAIM-) NAT INST IMMUNOLOGY.
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX PI Lal AA, Shi YP, Hasnain SE;
XX DR WPI: 2000-237654/20.
XX DR N-PSDB; RAZ51336.
XX
XX Novel recombinant protein as vaccine for treating malarial infection
XX PT comprises antigenic peptides obtained from different stages of
XX PT plasmodium falciparum life cycle -
XX
XX Claim 3: Page 43-44; 52pp; English.
XX
XX The present sequence is that of recombinant protein CDC/NIIMALVAC-1,
XX CC which is a multivalent, multistage malarial vaccine. The recombinant
XX CC protein comprises, melittin signal peptide, (His)6 tag, T-cell epitope
XX CC from tetanus toxoid and 21 antigenic epitopes from circumsporozoite
XX CC protein (CSP), sporozoite surface protein-2 (SSP-2), liver stage
XX CC antigen-1 (LSA-1), merozoite surface protein-1 (MSP-1), MSP-2, apical
XX CC membrane antigen-1 (AMA-1), erythrocyte binding antigen-175 (EBA-175),
XX CC rhoptry associated protein-1 (RAP-1) and gamete specific antigen, Pf27.
XX CC These epitopes were obtained at different stages of the life cycle of
XX CC Plasmodium falciparum. CDC/NIIMALVAC-1 vaccine has antiparasitic
XX CC activity and can be used for treatment and prevention of malarial
XX CC infections. Anti-CDC/NIIMALVAC-1 antibodies can be used for detecting
XX CC P. falciparum in biological samples.
XX
XX Sequence 350 AA:
XX
XX Query Match 100.0%; Score 84; DB 21; Length 350;
XX Best Local Similarity 100.0%; Pred. No. 3.2e-06;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 LATRLMKKFKAEIRDFF 17
XX Db 317 latrlmkkfkaeirdff 333
XX
XX RESULT 3
XX ID AAR06991 standard; protein: 782 AA.
XX AC AAR06991;
XX
XX DT 16-JAN-1991 (first entry)
XX DE Polypeptide antigenic to rhostry organelles of Plasmodium.
XX KW Malaria; vaccine; rhostry organelles..
XX OS Plasmodium falciparum.
XX
XX PN EP388738-A.
XX
XX PD 26-SEP-1990.
XX
XX PF 09-MAR-1990; 90EP-0104561.
XX
XX PR 22-AUG-1989; 89GB-0019064.
XX PR 14-MAR-1989; 89GB-0005857.
XX
XX PA (HOFF ) HOFFMANN-LA ROCHE AG.
XX PI Ridley RG, Scaife JG;
XX DR WPI: 1990-291721/39.
XX DR N-PSDB; AAO06000.
XX
XX Antigenic polypeptide and DNA encoding it - having a determinant
XX PT cross reactive with those on the rhostry organelles of the

```



PT merozoite form of the malaria parasite P falciparum.

PS Claim 1: Fig 2a-b; 29pp; English.

CC Gene product may be isolated from a transformed E. coli (Y1088) expression system using plasmid pMC9, and may then be used for CC immunisation against malaria. The product may also be used CC diagnostically to detect Abs directed against the parasite.

XX Sequence 782 AA;

Query Match 100.0%; Score 84; DB 11; Length 782;

Best Local Similarity 100.0%; Pred. NO. 7.5e-06;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LATRLMKKFKAEIRDF 17

Db 736 latrimkkfkaeirdf 752

RESULT 4

AAG82568

ID AAG82568 standard; Protein; 190 AA.

XX AC AAG82568;

DT 03-SEP-2001 (first entry)

DE S. epidermidis open reading frame protein sequence SEQ ID NO:2230.

XX Staphylococcus epidermidis SRI strain; infection; diagnosis;

KW vaccination; endocarditis.

XX Staphylococcus epidermidis.

OS WO200134809-A2.

XX 17-MAY-2001.

XX 09-NOV-2000; 2000WO-US30782.

XX 09-NOV-1999; 99US-0164258.

XX (GLAX ) GLAXO GROUP LTD.

XX Kimmerly WJ;

XX WPI; 2001-316495/33.

XX N-PSDB; AAH53418.

PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis, useful for vaccinating against infections, e.g. endocarditis -

XX Claim 18; Page 597; 2188pp; English.

XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.

CC (I) and (II) can have antibacterial activity and therefore can be used CC in vaccination. The nucleic acids (I) may be used to produce the CC S. epidermidis polypeptides (II) via the production of vectors CC containing them which are used to produce hosts cells which express the CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be CC used to vaccinate subjects and to raise antibodies against the bacteria.

CC The polypeptides may also be used to assay for other inhibitors of their CC activity and therefore identify compounds that may be used for the CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to CC AAH5090 represent specifically claimed S. epidermidis genomic DNA CC polynucleotide sequences from the present invention. AAH55091 to CC AAH55098 represent oligonucleotide sequences and primers which are used CC N.B. The present invention specifically claims all the polynucleotide CC sequences given in the sequence listing of the present specification.

CC however the sequence listing only goes up to SEQ ID NO:4454 so even CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472, CC no sequences are present for SEQ ID NO:4455 to 4464.

XX Sequence 190 AA;

Query Match 53.6%; Score 45; DB 22; Length 190;

Best Local Similarity 52.9%; Pred. NO. 5.3;

Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 LATRLMKKFKAEIRDF 17

Db 120 lmtamikkklngikdf 136

RESULT 5

AAG82906

ID AAG82906 standard; Protein; 1572 AA.

XX AC AAG82906;

DT 03-SEP-2001 (first entry)

DE S. epidermidis open reading frame protein sequence SEQ ID NO:2906.

XX Staphylococcus epidermidis SRI strain; infection; diagnosis;

KW vaccination; endocarditis.

XX Staphylococcus epidermidis.

OS WO200134809-A2.

XX 17-MAY-2001.

XX 09-NOV-2000; 2000WO-US30782.

XX 09-NOV-1999; 99US-0164258.

XX (GLAX ) GLAXO GROUP LTD.

XX Kimmerly WJ;

XX WPI; 2001-316495/33.

XX N-PSDB; AAH53756.

PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis, useful for vaccinating against infections, e.g. endocarditis -

XX Claim 18; Page 759; 2188pp; English.

XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.

CC (I) and (II) can have antibacterial activity and therefore can be used CC in vaccination. The nucleic acids (I) may be used to produce the CC S. epidermidis polypeptides (II) via the production of vectors CC containing them which are used to produce hosts cells which express the CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be CC used to vaccinate subjects and to raise antibodies against the bacteria.

CC The polypeptides may also be used to assay for other inhibitors of their CC activity and therefore identify compounds that may be used for the CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to CC AAH5090 represent specifically claimed S. epidermidis genomic DNA CC polynucleotide sequences from the present invention. AAH55091 to CC AAH55098 represent oligonucleotide sequences and primers which are used CC N.B. The present invention specifically claims all the polynucleotide CC sequences given in the sequence listing of the present specification,

CC however the sequence listing only goes up to SEQ ID NO:4454 so even CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472, CC no sequences are present for SEQ ID NO:4455 to 4464.

XX Sequence 1572 AA;

polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. AAC79790 to AAC79798 and AAB44761 represent sequences used in the exemplification of the present invention.

Query Match 51.2%; Score 43; DB 21; Length 86;  
Best Local Similarity 53.3%; Pred. No. 5;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Sequence 86 AA;

1 LATRLMKKFKAEIRD 15  
||||:||||:  
67 latrfkksfkanlen 81

RESULT 7  
AAB44823  
ID AAB44823 standard; Protein; 86 AA.  
XX AC AAB44823;  
XX AC AAB44823;  
DT 12-FEB-2001 (first entry)  
XX DE Human secreted protein sequence encoded by gene 36 SEQ ID NO:122.  
XX KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;  
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;  
KW cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;  
KW fungicide; ophthalmological; gene therapy; autoimmune disease; infection;  
KW hyperproliferative disorder; cardiovascular disorder; angiogenesis;  
KW cerebrovascular disorder; nervous system disorder; ocular disorder;  
KW wound healing; skin aging; food additive; preservative.  
XX OS Homo sapiens.  
XX OS WO200058336-A1.  
XX PN 05-OCT-2000.  
XX PD 23-MAR-2000; 2000WO-US07726.  
XX PF 26-MAR-1999; 99US-0126597.  
XX PR 07-JAN-2000; 2000US-0174877.  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX PI Rosen CA, Ruben SM, Komatsoulis G;  
XX WPI; 2000-602355/57.  
XX Nucleic acid encoding human secreted proteins, used to treat, prevent, ameliorate or diagnose medical conditions such as cancer, and autoimmune diseases -  
XX Disclosure; Page 381; 391pp; English.  
XX The polynucleotide sequences given in AAC79799 to AAC79848 encode the human secreted proteins given in AAB44762 to AAB44811. AAB44812 to AAB44829 represent human secreted polypeptide sequences and proteins homologous to them, which are used in the exemplification of the present invention. Human secreted proteins have activities based on the tissues and cells the genes are expressed in. Examples of activities are: immunosuppressive; antiarthritic; antirheumatic; antiproliferative; neuroprotective; antibacterial; virucide; fungicide; and ophthalmological. The polynucleotides and polypeptides can be used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Disorders which are diagnosed or treated include autoimmune diseases, hyperproliferative disorders, cardiovascular

polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. AAC79790 to AAC79798 and AAB44761 represent sequences used in the exemplification of the present invention.

Query Match 53.6%; Score 45; DB 22; Length 1572;  
Best Local Similarity 52.9%; Pred. No. 48;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

1 LATRLMKKFKAEIRDPF 17  
||:|||||:  
1161 lmtamikkknkngikdff 1177

RESULT 6  
AAB44822  
ID AAB44822 standard; Protein; 86 AA.  
XX AC AAB44822;  
XX AC AAB44822;  
DT 12-FEB-2001 (first entry)  
XX DE Gene 36 human secreted protein homologous amino acid sequence #121.  
XX KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;  
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;  
KW cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;  
KW fungicide; ophthalmological; gene therapy; autoimmune disease; infection;  
KW hyperproliferative disorder; cardiovascular disorder; angiogenesis;  
KW cerebrovascular disorder; nervous system disorder; ocular disorder;  
KW wound healing; skin aging; food additive; preservative.  
XX OS Homo sapiens.  
XX OS WO200058336-A1.  
XX PN 05-OCT-2000.  
XX PD 23-MAR-2000; 2000WO-US07726.  
XX PF 26-MAR-1999; 99US-0126597.  
XX PR 07-JAN-2000; 2000US-0174877.  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX PI Rosen CA, Ruben SM, Komatsoulis G;  
XX WPI; 2000-602355/57.  
XX Nucleic acid encoding human secreted proteins, used to treat, prevent, ameliorate or diagnose medical conditions such as cancer, and autoimmune diseases -  
XX Disclosure; Page 380-381; 391pp; English.  
XX The polynucleotide sequences given in AAC79799 to AAC79848 encode the human secreted proteins given in AAB44762 to AAB44811. AAB44812 to AAB44829 represent human secreted polypeptide sequences and proteins homologous to them, which are used in the exemplification of the present invention. Human secreted proteins have activities based on the tissues and cells the genes are expressed in. Examples of activities are: immunosuppressive; antiarthritic; antirheumatic; antiproliferative; neuroprotective; antibacterial; virucide; fungicide; and ophthalmological. The polynucleotides and polypeptides can be used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Disorders which are diagnosed or treated include autoimmune diseases, hyperproliferative disorders, cardiovascular disorders, cerebrovascular disorders, angiogenesis, nervous system disorders, infections caused by bacteria, viruses and fungi and ocular disorders. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The

CC disorders, cerebrovascular disorders, angiogenesis, nervous system  
CC disorders, infections caused by bacteria, viruses and fungi and ocular  
CC disorders. The polypeptides can also be used to aid wound healing and  
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to  
CC maintain organs before transplantation, for supporting cell culture of  
CC primary tissues, to regenerate tissues and in chemotaxis. The  
CC polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities. AAC79790 to AAC79798 and  
CC AAB44761 represent sequences used in the exemplification of the present  
XX invention.  
SQ Sequence 86 AA;

Query Match 51.2%; Score 43; DB 21; Length 86;  
Best Local Similarity 53.3%; Pred. No. 5;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 LATRLMKKFKAEIRD 15  
||||:| ||| : :  
Db 67 latrflksfkanlen 81

RESULT 8  
AAB57133  
ID AAB57133 standard; Protein; 98 AA.

XX AAB57133;

XX 13-MAR-2001 (first entry)

DE Human prostate cancer antigen protein sequence SEQ ID NO:1711.

XX Human; prostate cancer; prostate cancer antigen; detection; diagnosis;  
KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;  
KW vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;  
KW antibacterial; gene therapy; neural; immune; reproductive; renal;  
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;  
KW wound; infectious disease.

XX Homo sapiens.

XX WO200055174-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05988.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.  
PA (ROSE/) ROSEN C A.

XX Rosen CA, Ruben SM;

XX WPI; 2000-587513/55.

XX N-PSDB; AAF16336.

PT Prostate cancer associated gene sequences, referred to as prostate  
PT cancer antigens, useful for treatment, prevention, and diagnosis of  
PT disorders such as prostate cancer.

XX Claim 11; Page 2184; 2338pp; English.

XX AAF15566 to AAF16505 encode the human prostate cancer associated  
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.  
CC The prostate cancer antigens can have neuroprotective, cytostatic,  
CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,  
CC nephrotropic, antiinfective, gynaecological and antibacterial activities,  
CC and can be used in gene therapy. The prostate cancer antigen  
CC polynucleotides may be used for detection of prostate cancer, chromosome  
CC identification, as chromosome markers, and for numerous other diagnostic  
CC or research purposes. The prostate cancer antigens may be used to treat

CC disorders such as neural, immune, muscular, reproductive,  
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to  
CC AAB57303 represent sequences used in the exemplification of the present  
CC invention.  
XX Sequence 98 AA;

Query Match 51.2%; Score 43; DB 21; Length 98;  
Best Local Similarity 53.3%; Pred. No. 5.7;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 LATRLMKKFKAEIRD 15  
||||:| ||| : :  
Db 79 latrflksfkanlen 93

RESULT 9  
AAG31171  
ID AAG31171 standard; Protein; 297 AA.

XX AAG31171;

XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 37389.

XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 28-APR-1999; 99US-0130891.

XX 30-APR-1999; 99US-0131449.

XX 30-APR-1999; 99US-0132048.

XX 04-MAY-1999; 99US-0132407.

XX 05-MAY-1999; 99US-0132484.

XX 06-MAY-1999; 99US-0132485.

XX 06-MAY-1999; 99US-0132486.

XX 07-MAY-1999; 99US-0132863.

XX 11-MAY-1999; 99US-0134256.

XX 14-MAY-1999; 99US-0134218.

XX 14-MAY-1999; 99US-0134219.

XX 14-MAY-1999; 99US-0134221.

XX 18-MAY-1999; 99US-0134370.

XX 18-MAY-1999; 99US-0134768.

XX 19-MAY-1999; 99US-0134941.

XX 20-MAY-1999; 99US-0135124.

XX 21-MAY-1999; 99US-0135353.

XX 24-MAY-1999; 99US-0135629.

XX 25-MAY-1999; 99US-0136021.

us-09-763-397a-23.rag

Mon Feb 4 15:23:44 2002

PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140354.  
PR 28-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143342.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 05-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 15-SEP-1999; 99US-0154039.  
PR 16-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 28-OCT-1999; 99US-0161993.

```
PR 29-OCT-1999; 99US-0162142.
Query Match 51.2%; Score 43; DB 21; Length 297;
Best Local Similarity 58.3%; Pred. No. 18;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
QY 6 MKKFAEIRDF 17
: |||::|||
Db 36 ldkfsdlqdf 47

RESULT 10
AAG311170
ID AAG311170 standard; Protein; 335 AA.
XX
AC AAG311170;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37388.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
PD
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140895.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
```

PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150560.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	14-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161922.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match 51.2%; Score 43; DB 21; Length 335;		
Best Local Similarity 58.3%; Pred. No. 21;		
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;		
6 MKKFAEIRDEF 17		
QY		
Db	74	idkfsdldgff 85
:   :: : :		
RESULT 11		
AAG31169		
ID	AAG31169	standard; Protein; 353 AA.
XX		
AC	AAG31169;	
XX		
DT	17-OCT-2000	(first entry)
XX		
DE	Arabidopsis thaliana	protein fragment SEQ ID NO: 37387.
XX		
KW	Protein identification; signal transduction pathway; metabolic pathway;	
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	
KW	termination sequence.	
XX		
OS	Arabidopsis thaliana.	
XX		
PN	EP1033405-A2.	
XX		
PD	06-SEP-2000.	
XX		
PF	25-FEB-2000;	2000EP-0301439.
XX		
XX	25-FEB-1999;	99US-0121825.
PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	28-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.

PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140891.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.

PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 06-OCT-1999; 99US-0157753.  
PR 07-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 51.2%; Score 43; DB 21; Length 353;

Best Local Similarity 58.3%; Pred. No. 22;

Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 6 MKFKAEIRDF 17

: |||:|:|

Db 92 ldkfsdiqdf 103

RESULT 12

AAG04103

ID AAG04103 standard; Protein; 186 AA.

XX AC AAG04103;  
XX DT 17-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 4.  
XX KW protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139464.  
PR 18-JUN-1999; 99US-0139465.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140354.  
PR 23-JUN-1999; 99US-0140695.  
PR 24-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140991.  
PR 29-JUN-1999; 99US-0141287.  
PR 30-JUN-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.



```
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      50.0%; Score 42; DB 21; Length 186;
Best Local Similarity 69.2%; Pred. No. 16;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 5 LMKKFAEIRDF 17
Db 10 lmkvqgeirdrf 22

RESULT 13
AAG04102
ID AAG04102 standard; Protein; 190 AA.
XX
AC AAG04102;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 3.
XX
```

```
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
```

PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 20-JUL-1999; 99US-0144684.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 23-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145313.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 23-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.

PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 22-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 50.0%; Score 42; DS 21; Length 190;  
 Best Local Similarity 69.2%; Pred. No. 17;  
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 5 LMKFKAEIRDF 17  
 Db 14 lmkvgeirdrf 26  
 |||| : |||| |

RESULT 14

AAG04101  
 ID AAG04101 standard; Protein; 203 AA.

XX  
 AC AAG04101;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 2.

XX Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

PN EF1033405-A2.

XX 06-SEP-2000.  
PD 25-FEB-2000; 2000EP-0301439.  
XX  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135622.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137503.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.

Mon Feb 4 15:23:44 2002

PR	22-SEP-1999;	99US-0155139.	PR	09-MAR-1999;	99US-0123548.
PR	23-SEP-1999;	99US-0155486.	PR	23-MAR-1999;	99US-0125788.
PR	24-SEP-1999;	99US-0155659.	PR	25-MAR-1999;	99US-0126264.
PR	28-SEP-1999;	99US-0156458.	PR	29-MAR-1999;	99US-0126785.
PR	29-SEP-1999;	99US-0156596.	PR	01-APR-1999;	99US-0127462.
PR	04-OCT-1999;	99US-0157117.	PR	06-APR-1999;	99US-0128234.
PR	05-OCT-1999;	99US-0157753.	PR	08-APR-1999;	99US-0128714.
PR	06-OCT-1999;	99US-01577865.	PR	16-APR-1999;	99US-0129845.
PR	07-OCT-1999;	99US-0158029.	PR	19-APR-1999;	99US-0130077.
PR	08-OCT-1999;	99US-0158232.	PR	21-APR-1999;	99US-0130449.
PR	12-OCT-1999;	99US-0158369.	PR	23-APR-1999;	99US-0130510.
PR	13-OCT-1999;	99US-0159293.	PR	23-APR-1999;	99US-0130891.
PR	13-OCT-1999;	99US-0159294.	PR	28-APR-1999;	99US-0131449.
PR	13-OCT-1999;	99US-0159295.	PR	30-APR-1999;	99US-0132048.
PR	14-OCT-1999;	99US-0159329.	PR	04-MAY-1999;	99US-0132407.
PR	14-OCT-1999;	99US-0159330.	PR	05-MAY-1999;	99US-0132484.
PR	14-OCT-1999;	99US-0159331.	PR	05-MAY-1999;	99US-0132485.
PR	14-OCT-1999;	99US-0159637.	PR	06-MAY-1999;	99US-0132486.
PR	14-OCT-1999;	99US-0159638.	PR	06-MAY-1999;	99US-0132487.
PR	18-OCT-1999;	99US-0159384.	PR	07-MAY-1999;	99US-0132863.
PR	21-OCT-1999;	99US-0160741.	PR	11-MAY-1999;	99US-0134256.
PR	21-OCT-1999;	99US-0160767.	PR	11-MAY-1999;	99US-0134218.
PR	21-OCT-1999;	99US-0160768.	PR	14-MAY-1999;	99US-0134219.
PR	21-OCT-1999;	99US-0160770.	PR	14-MAY-1999;	99US-0134221.
PR	21-OCT-1999;	99US-0160814.	PR	14-MAY-1999;	99US-0134370.
PR	21-OCT-1999;	99US-0160815.	PR	18-MAY-1999;	99US-0134768.
PR	22-OCT-1999;	99US-0160980.	PR	19-MAY-1999;	99US-0134941.
PR	22-OCT-1999;	99US-0160981.	PR	20-MAY-1999;	99US-0135124.
PR	22-OCT-1999;	99US-0160989.	PR	21-MAY-1999;	99US-0135353.
PR	25-OCT-1999;	99US-0161404.	PR	24-MAY-1999;	99US-0135629.
PR	25-OCT-1999;	99US-0161405.	PR	25-MAY-1999;	99US-0136021.
PR	25-OCT-1999;	99US-0161406.	PR	27-MAY-1999;	99US-0136392.
PR	26-OCT-1999;	99US-0161359.	PR	28-MAY-1999;	99US-0136782.
PR	26-OCT-1999;	99US-0161360.	PR	01-JUN-1999;	99US-0137222.
PR	26-OCT-1999;	99US-0161361.	PR	03-JUN-1999;	99US-0137528.
PR	28-OCT-1999;	99US-0161920.	PR	04-JUN-1999;	99US-0137502.
PR	28-OCT-1999;	99US-0161992.	PR	07-JUN-1999;	99US-0137724.
PR	28-OCT-1999;	99US-0161993.	PR	08-JUN-1999;	99US-0138094.
PR	28-OCT-1999;	99US-0161993.	PR	10-JUN-1999;	99US-0138540.
PR	29-OCT-1999;	99US-0162142.	PR	10-JUN-1999;	99US-0138847.
PR	29-OCT-1999;	99US-0162142.	PR	14-JUN-1999;	99US-0139119.
PR	29-OCT-1999;	99US-0162142.	PR	16-JUN-1999;	99US-0139452.
PR	29-OCT-1999;	99US-0162142.	PR	16-JUN-1999;	99US-0139453.
PR	29-OCT-1999;	99US-0162142.	PR	17-JUN-1999;	99US-0139492.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139454.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139455.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139456.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139457.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139458.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139459.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139460.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139461.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139462.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139463.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139750.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139763.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139817.
PR	29-OCT-1999;	99US-0162142.	PR	21-JUN-1999;	99US-0139817.
PR	29-OCT-1999;	99US-0162142.	PR	22-JUN-1999;	99US-0139899.
PR	29-OCT-1999;	99US-0162142.	PR	23-JUN-1999;	99US-0140353.
PR	29-OCT-1999;	99US-0162142.	PR	23-JUN-1999;	99US-0140354.
PR	29-OCT-1999;	99US-0162142.	PR	24-JUN-1999;	99US-0140695.
PR	29-OCT-1999;	99US-0162142.	PR	28-JUN-1999;	99US-0140823.
PR	29-OCT-1999;	99US-0162142.	PR	29-JUN-1999;	99US-0140991.
PR	29-OCT-1999;	99US-0162142.	PR	30-JUN-1999;	99US-0141287.
PR	29-OCT-1999;	99US-0162142.	PR	01-JUL-1999;	99US-0141842.
PR	29-OCT-1999;	99US-0162142.	PR	01-JUL-1999;	99US-0142154.
PR	29-OCT-1999;	99US-0162142.	PR	02-JUL-1999;	99US-0142055.
PR	29-OCT-1999;	99US-0162142.	PR	06-JUL-1999;	99US-0142390.
PR	29-OCT-1999;	99US-0162142.	PR	08-JUL-1999;	99US-0142803.
PR	29-OCT-1999;	99US-0162142.	PR	09-JUL-1999;	99US-0142920.
PR	29-OCT-1999;	99US-0162142.	PR	12-JUL-1999;	99US-0142977.
PR	29-OCT-1999;	99US-0162142.	PR	13-JUL-1999;	99US-0143542.
PR	29-OCT-1999;	99US-0162142.	PR	14-JUL-1999;	99US-0143624.

Query Match 50.0%; Score 42; DB 21; Length 203;  
Best Local Similarity 69.2%; Pred. No. 18;  
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 5 LMKKFAEIRDF 17  
|||||:|||||  
Db 27 lmkkvqgeirdrf 39

RESULT 15  
AAG07741  
ID AAG07741 standard; Protein; 300 AA.  
XX  
AC AAG07741;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 5008.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
FD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.

PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.

PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160983.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 48.8%; Score 41; DB 21; Length 300;  
Best Local Similarity 50.0%; Pred. No. 40;  
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;  
Qy 1 LATRLMKKFAEIRDF 16  
|||:| | |||  
Db 276 laqriqekikgnyrdf 291

Search completed: January 29, 2002, 10:59:19  
Job time: 2243 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 10:59:57 ; Search time 32.24 Seconds  
(without alignments)  
11.866 Million cell updates/sec

Title: US-09-763-397A-23

Perfect score: 84

Sequence: 1 LATRLMKFKAEIRDF 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

- 1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	DB ID	Description
1	41	48.8	2325	3	US-08-417-089-6
2	41	48.8	2325	4	US-08-695-651-6
3	41	48.8	2325	4	US-08-930-285-6
4	41	48.8	2325	4	US-08-695-421-6
5	40	47.6	775	1	US-07-603-133B-15
6	40	47.6	775	1	US-07-603-133B-16
7	39	46.4	1147	3	US-08-470-260-5
8	39	46.4	1147	3	US-08-471-491-5
9	39	46.4	1147	4	US-08-466-662-5
10	39	46.4	3289	2	US-08-477-451-2
11	37	44.0	306	1	US-08-089-986-2
12	37	44.0	306	1	US-08-478-585-2
13	37	44.0	306	1	US-08-717-312-2
14	37	44.0	306	2	US-08-266-408-2
15	37	44.0	306	5	PCT-US94-07886-2
16	37	44.0	776	1	US-07-603-133B-18
17	37	44.0	859	1	US-08-053-614-2
18	37	44.0	859	1	US-08-316-397B-2
19	37	44.0	859	2	US-09-034-306-2
20	37	44.0	859	4	US-09-259-437-2
21	37	44.0	859	5	PCT-US93-09782-2
22	37	44.0	1181	1	US-08-053-614-4
23	37	44.0	1181	1	US-08-316-397B-4
24	37	44.0	1181	2	US-09-034-306-4
25	37	44.0	1181	4	US-09-259-437-4
26	37	44.0	1181	5	PCT-US93-09782-4
27	37	44.0	1784	4	US-09-040-738-2

28	37	44.0	1784	4	US-08-652-426A-2	Sequence 2, Appl
29	35	41.7	20	1	US-08-311-611A-57	Sequence 57, Appl
30	35	41.7	20	1	US-08-311-611A-150	Sequence 150, App
31	35	41.7	20	1	US-08-372-783-57	Sequence 57, Appl
32	35	41.7	20	1	US-08-372-783-150	Sequence 150, App
33	35	41.7	20	1	US-08-372-105-57	Sequence 57, Appl
34	35	41.7	20	1	US-08-372-105-150	Sequence 150, App
35	35	41.7	20	1	US-08-306-473A-57	Sequence 57, Appl
36	35	41.7	20	1	US-08-306-473A-150	Sequence 150, App
37	35	41.7	20	1	US-08-209-762-57	Sequence 57, Appl
38	35	41.7	20	2	US-08-621-803-47	Sequence 47, Appl
39	35	41.7	20	2	US-08-485-445A-57	Sequence 57, Appl
40	35	41.7	20	2	US-08-485-445A-150	Sequence 150, App
41	35	41.7	20	3	US-09-119-263-57	Sequence 57, Appl
42	35	41.7	20	3	US-09-119-263-150	Sequence 150, App
43	35	41.7	20	4	US-08-657-162-57	Sequence 57, Appl
44	35	41.7	20	4	US-08-657-162-150	Sequence 150, App
45	35	41.7	20	4	US-09-224-480-57	Sequence 57, Appl

ALIGNMENTS

RESULT 1  
US-08-417-089-6  
; Sequence 6, Application US/08417089  
; Patent No. 6069298  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: METHODS AND AN ACETYL COA CARBOXYLASE GENE  
; TITLE OF INVENTION: FOR CONFERRING HERBICIDE TOLERANCE AND AN ALTERATION IN  
; TITLE OF INVENTION: OIL CONTENT OF PLANTS  
; NUMBER OF SEQUENCES: 11  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION NUMBER: US/08/417,089  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2325 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-417-089-6

Query Match 48.8%; Score 41; DB 3; Length 2325;  
Best Local Similarity 53.3%; Pred. No. 1.1e+02;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 LATRLMKFKAEIRD 15  
| | | | | : | : | : |  
Db 919 LATRLPRNLKSELED 933

RESULT 2  
US-08-695-651-6  
; Sequence 6, Application US/08695651  
; Patent No. 6146867  
; GENERAL INFORMATION:  
; APPLICANT: Gengenbach, B. G.  
; APPLICANT: Somers, D. A.  
; APPLICANT: Wyse, D. L.  
; APPLICANT: Gronwald, J. W.  
; APPLICANT: Egli, M. A.  
; APPLICANT: Lutz, S. M.  
; TITLE OF INVENTION: METHOD FOR PRODUCING PLANT ACETYL COA  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P. A.  
STREET: P.O. Box 2938  
CITY: Minneapolis  
STATE: MN  
COUNTRY: USA  
ZIP: 55402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/695,651  
FILING DATE: 12-AUG-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/417089  
FILING DATE: 05-APR-1995  
APPLICATION NUMBER: 08/014326  
FILING DATE: 05-FEB-1993  
APPLICATION NUMBER: 07/917462  
FILING DATE: 21-JUL-1992  
APPLICATION NUMBER: 07/538674  
FILING DATE: 18-JUN-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Woessner, Warren D  
REGISTRATION NUMBER: 30,440  
REFERENCE/DOCKET NUMBER: 600.318US3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-373-6900  
TELEFAX: 612-339-3061  
TELEX:

INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2325 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: N-terminal  
US-08-695-651-6

Query Match 48.8%; Score 41; DB 4; Length 2325;  
Best Local Similarity 53.3%; Pred. No. 1.1e+02;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 LATRLMKKFAEIRD 15  
||||| : ||: |

Db 919 LATRLPRNLKSELED 933

## RESULT 3

US-08-930-285-6  
Sequence 6, Application US/08930285  
Patent No. 6222099  
GENERAL INFORMATION:

APPLICANT: Regents of the University of Minnesota, et al.  
TITLE OF INVENTION: TRANSGENIC PLANTS EXPRESSING ACETYL COA CARBOXYLASE GE  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P. A.  
STREET: P. O. Box 2938  
CITY: Minneapolis  
STATE: MN  
COUNTRY: USA  
ZIP: 55402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/930,285  
FILING DATE: 13-APR-1998  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/04625  
FILING DATE: 04-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Embretson, Janet E.  
REGISTRATION NUMBER: 39,665  
REFERENCE/DOCKET NUMBER: 600.318US4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-339-0331  
TELEFAX: 612-339-3061  
TELEX:  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2325 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal and C-terminal (full length protein)  
ORIGINAL SOURCE:  
US-08-930-285-6

Query Match 48.8%; Score 41; DB 4; Length 2325;  
Best Local Similarity 53.3%; Pred. No. 1.1e+02;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 LATRLMKKFAEIRD 15

Db 919 LATRLPRNLKSELED 933

## RESULT 4

US-08-695-421-6  
Sequence 6, Application US/08695421  
Patent No. 6268550  
GENERAL INFORMATION:

APPLICANT: Gengenbach, B. G.  
APPLICANT: Somers, D. A.  
APPLICANT: Wyse, D. L.  
APPLICANT: Gronwald, J. W.  
APPLICANT: Egli, M. A.  
APPLICANT: Lutz, S. M.  
TITLE OF INVENTION: METHODS AND AN ACETYL COA CARBOXYLASE GENE  
TITLE OF INVENTION: FOR CONFERRING HERBICIDE TOLERANCE AND AN ALTERATION IN OIL  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P. A.  
STREET: P.O. Box 2938  
CITY: Minneapolis  
STATE: MN  
COUNTRY: USA  
ZIP: 55402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/695,421  
FILING DATE: 23-AUG-1996  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/417089  
FILING DATE: 05-APR-1995  
APPLICATION NUMBER: 08/014326  
FILING DATE: 05-FEB-1993  
APPLICATION NUMBER: 07/917462



;; FILING DATE: 21-JUL-1992  
;; APPLICATION NUMBER: 07/538674  
;; FILING DATE: 18-JUN-1990  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Woessner, Warren D  
;; REGISTRATION NUMBER: 30,440  
;; REFERENCE/DOCKET NUMBER: 600.318US2  
;; TELEPHONE: 612-373-6900  
;; TELEFAX: 612-339-3061  
;; TELEX:  
;; INFORMATION FOR SEQ ID NO: 6:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 2325 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; FRAGMENT TYPE: N-terminal  
;; US-08-695-421-6

Query Match 48.8%; Score 41; DB 4; Length 2325;  
Best Local Similarity 53.3%; Pred. No. 1.1e+02;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 LATRLMKKFAEIRD 15  
      ||||| : 1:1: |  
Db 919 LATRLPRNLKSELD 933

RESULT 5  
US-07-603-133B-15  
; Sequence 15, Application US/07603133B  
; Patent No. 5298244  
; GENERAL INFORMATION:  
; APPLICANT: Redmond, Mark J.  
; APPLICANT: Ijaz, Mohammed K.  
; TITLE OF INVENTION: ASSEMBLED VIRAL PARTICLES AND THEIR  
; TITLE OF INVENTION: USE IN A VACCINE TO ROTAVIRAL DISEASE  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Morrison & Foerster  
; STREET: 545 Middlefield Road, Suite 200  
; CITY: Menlo Park  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94025  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; FILING DATE: 19901025  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Robins, Roberta L.  
; REGISTRATION NUMBER: 33,208  
; REFERENCE/DOCKET NUMBER: 9313-0004.00  
; TELEPHONE: (415) 327-7250  
; TELEFAX: (415) 327-2951  
; TELEX: 706141  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 775 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein

US-07-603-133B-15

Query Match 47.6%; Score 40; DB 1; Length 775;  
Best Local Similarity 70.0%; Pred. No. 55;  
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LATRLMKKFK 10  
      :|||||  
Db 543 MATKVMKKFK 552

RESULT 6  
US-07-603-133B-16  
; Sequence 16, Application US/07603133B  
; Patent No. 5298244  
; GENERAL INFORMATION:  
; APPLICANT: Redmond, Mark J.  
; APPLICANT: Ijaz, Mohammed K.  
; APPLICANT: Parker, Michael D.  
; TITLE OF INVENTION: ASSEMBLED VIRAL PARTICLES AND THEIR  
; TITLE OF INVENTION: USE IN A VACCINE TO ROTAVIRAL DISEASE  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Morrison & Foerster  
; STREET: 545 Middlefield Road, Suite 200  
; CITY: Menlo Park  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94025  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC-compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/603.133B  
; FILING DATE: 19901025  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Robins, Roberta L.  
; REGISTRATION NUMBER: 33,208  
; REFERENCE/DOCKET NUMBER: 9313-0004.00  
; TELEPHONE: (415) 327-7250  
; TELEFAX: (415) 327-2951  
; TELEX: 706141  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 775 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; US-07-603-133B-16

Query Match 47.6%; Score 40; DB 1; Length 775;  
Best Local Similarity 70.0%; Pred. No. 55;  
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LATRLMKKFK 10  
      :|||||  
Db 543 MATKVMKKFK 552

RESULT 7  
US-08-470-260-5  
; Sequence 5, Application US/08470260  
; Patent No. 6077706  
; GENERAL INFORMATION:  
; APPLICANT: Covacci, Antonello  
; APPLICANT: Bugnoli, Massimo

APPLICANT: Telford, John  
APPLICANT: Macchia, Giovanni  
APPLICANT: Rappuoli, Rino  
TITLE OF INVENTION: Helicobacter Pylori Proteins Useful  
for Vaccines and Diagnostics  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Chiron Corporation  
STREET: 4560 Horton Street  
CITY: Emeryville  
STATE: California  
COUNTRY: USA  
ZIP: 94608-2916  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/470,260  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/256,848  
FILING DATE: 21-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: McClung, Barbara G.  
REGISTRATION NUMBER: 33,113  
REFERENCE/DOCKET NUMBER: 0316.001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 801-2708  
TELEFAX: (510) 653-3342  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1147 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-470-260-5

Query Match 46.4%; Score 39; DB 3; Length 1147;  
Best Local Similarity 43.8%; Pred. No. 1.2e+02;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 LATRLMKKFAEIRDF 16  
| : : ||: ||: ||

DB 413 LSEKEKEKFRTEIKDF 428

RESULT 8  
US-08-471-491-5  
Sequence 5, Application US/08471491B  
Patent No. 6090611  
GENERAL INFORMATION:  
APPLICANT: Covacci, Antonello  
APPLICANT: Bugnoli, Massimo  
APPLICANT: Telford, John  
APPLICANT: Macchia, Giovanni  
APPLICANT: Rappuoli, Rino  
TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And  
FILE OF INVENTION: Diagnostics  
FILE REFERENCE: CHIR0044  
CURRENT APPLICATION NUMBER: US/08/471,491B  
CURRENT FILING DATE: 1995-06-06  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 5  
LENGTH: 1147  
TYPE: PRT  
ORGANISM: Helicobacter pylori  
US-08-471-491-5

Query Match 46.4%; Score 39; DB 3; Length 1147;  
Best Local Similarity 43.8%; Pred. No. 1.2e+02;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 LATRLMKKFAEIRDF 16  
| : : ||: ||: ||

DB 413 LSEKEKEKFRTEIKDF 428

RESULT 9  
US-08-466-662-5  
Sequence 5, Application US/08466662B  
Patent No. 6130059  
GENERAL INFORMATION:  
APPLICANT: Covacci, Antonello  
APPLICANT: Bugnoli, Massimo  
APPLICANT: Telford, John  
APPLICANT: Macchia, Giovanni  
APPLICANT: Rappuoli, Rino  
TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And  
FILE OF INVENTION: Diagnostics  
FILE REFERENCE: CHIR0057  
CURRENT APPLICATION NUMBER: US/08/466,662B  
CURRENT FILING DATE: 1995-06-06  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 5  
LENGTH: 1147  
TYPE: PRT  
ORGANISM: Helicobacter pylori  
US-08-466-662-5

Query Match 46.4%; Score 39; DB 4; Length 1147;  
Best Local Similarity 43.8%; Pred. No. 1.2e+02;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 LATRLMKKFAEIRDF 16  
| : : ||: ||: ||

DB 413 LSEKEKEKFRTEIKDF 428

RESULT 10  
US-08-477-451-2  
Sequence 2, Application US/08477451  
Patent No. 5928865  
GENERAL INFORMATION:  
APPLICANT: Covacci, Antonello  
TITLE OF INVENTION: Helicobacter Pylori Cagi Region  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Chiron Corporation  
STREET: 4560 Horton Street  
CITY: Emeryville  
STATE: CA  
COUNTRY: USA  
ZIP: 94608-2916  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,451  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: McClung, Barbara G.  
REGISTRATION NUMBER: 33,113  
REFERENCE/DOCKET NUMBER: 0335.002  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 510-601-2708  
TELEFAX: 510-655-3542  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3289 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-477-451-2

Query Match 46.4%; Score 39; DB 2; Length 3289;  
Best Local Similarity 43.8%; Pred. No. 3.3e+02;  
Matches 7; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LATRLMKKFKAEIRDF 16  
I: :||: ||:|  
Db 1944 LSEKEKFRTEIKDF 1959

## RESULT 11

US-08-089-986-2  
Sequence 2, Application US/08089986  
Patent No. 5441884

## GENERAL INFORMATION:

APPLICANT: Baum, James A.  
TITLE OF INVENTION: Bacillus thuringiensis Transposon Tn5401  
TITLE OF INVENTION: and its use in a Site-Specific Recombination System For  
TITLE OF INVENTION: Recombinant Bacillus thuringiensis Strain Development  
NUMBER OF SEQUENCES: 4

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Panitch Schwarze Jacobs & Nadel c/o A.S.  
STREET: 1601 Market Street, 36th Floor  
CITY: Philadelphia  
STATE: Pennsylvania  
COUNTRY: U.S.A.  
ZIP: 19103

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/089,986  
FILING DATE: 19930708

## CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Egoif, Christopher  
REGISTRATION NUMBER: 27633  
REFERENCE/DOCKET NUMBER: 7205-47  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-757-1590

## INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 306 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-089-986-2

Query Match 44.0%; Score 37; DB 1; Length 306;  
Best Local Similarity 31.2%; Pred. No. 68;  
Matches 5; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

Qy 1 LATRLMKKFKAEIRDF 16  
I: :||: ||:|  
Db 28 LHTKTLKEYTSLKDF 43

## RESULT 12

US-08-478-585-2  
Sequence 2, Application US/08478585  
Patent No. 5650308

## GENERAL INFORMATION:

APPLICANT: Baum, James A.  
TITLE OF INVENTION: Bacillus thuringiensis Transposon Tn5401  
TITLE OF INVENTION: and its use in a Site-Specific Recombination System For  
TITLE OF INVENTION: Recombinant Bacillus thuringiensis Strain Development  
NUMBER OF SEQUENCES: 4

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Panitch Schwarze Jacobs & Nadel c/o A.S.  
STREET: 1601 Market Street, 36th Floor  
CITY: Philadelphia  
STATE: Pennsylvania  
COUNTRY: U.S.A.  
ZIP: 19103

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/478,585

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/089,986

FILING DATE: 08-JUL-1993

ATTORNEY/AGENT INFORMATION:

NAME: Egoif, Christopher

REGISTRATION NUMBER: 27633

REFERENCE/DOCKET NUMBER: 7205-47

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-757-1590

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 306 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-478-585-2

Query Match 44.0%; Score 37; DB 1; Length 306;  
Best Local Similarity 31.2%; Pred. No. 68;  
Matches 5; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

Qy 1 LATRLMKKFKAEIRDF 16  
I: :||: ||:|  
Db 28 LHTKTLKEYTSLKDF 43

## RESULT 13

US-08-717-312-2

Sequence 2, Application US/08717312

Patent No. 5776449

## GENERAL INFORMATION:

APPLICANT: Baum, James A.

TITLE OF INVENTION: Bacillus thuringiensis Transposon Tn5401

TITLE OF INVENTION: and its use in a Site-Specific Recombination System For

TITLE OF INVENTION: Recombinant Bacillus thuringiensis Strain Development

NUMBER OF SEQUENCES: 4

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Panitch Schwarze Jacobs & Nadel c/o A.S. Nadel  
STREET: 1601 Market Street, 36th Floor  
CITY: Philadelphia  
STATE: Pennsylvania  
COUNTRY: U.S.A.  
ZIP: 19103

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

Mon Feb 4 15:23:45 2002

Db 28 LHTKTLKEYTSDLKDF 43

RESULT 15

PCT-US94-07886-2

Sequence 2, Application PC/TUS9407886

GENERAL INFORMATION:

APPLICANT: Baum, James A.

TITLE OF INVENTION: Bacillus thuringiensis Transposon Tn5401

TITLE OF INVENTION: and its use in a Site-Specific Recombination System For

TITLE OF INVENTION: Recombinant Bacillus thuringiensis Strain Development

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESS: Panitch Schwarze Jacobs & Nadel c/o A.S. Nadel

STREET: 1601 Market Street, 36th Floor

CITY: Philadelphia

STATE: Pennsylvania

COUNTRY: U.S.A.

ZIP: 19103

COMPUTER READABLE FORM: disk

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/07886

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/089,986

FILING DATE: 08-JUL-1993

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 306 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-US94-07886-2

Query Match 44.0%; Score 37; DB 5; Length 306;

Best Local Similarity 31.2%; Pred. No. 68;

Matches 5; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

QY 1 LATRLMKFKAEIRDF 16

Db 28 LHTKTLKEYTSDLKDF 43

Search completed: January 29, 2002, 10:59:58

Job time: 2144 sec

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/717,312

FILING DATE: 20-SEP-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/266,408

FILING DATE: 24-JUL-1994

APPLICATION NUMBER: US 08/089,986

FILING DATE: 08-JUL-1993

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 306 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-717-312-2

Query Match 44.0%; Score 37; DB 1; Length 306;

Best Local Similarity 31.2%; Pred. No. 68;

Matches 5; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

QY 1 LATRLMKFKAEIRDF 16

Db 28 LHTKTLKEYTSDLKDF 43

RESULT 14

US-08-266-408-2

Sequence 2, Application US/08266408

Patent No. 5843744

GENERAL INFORMATION:

APPLICANT: Baum, James A.

TITLE OF INVENTION: Bacillus thuringiensis Transposon Tn5401

TITLE OF INVENTION: and its use in a Site-Specific Recombination System For

TITLE OF INVENTION: Recombinant Bacillus thuringiensis Strain Development

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESS: Panitch Schwarze Jacobs & Nadel c/o A.S. Nadel

STREET: 1601 Market Street, 36th Floor

CITY: Philadelphia

STATE: Pennsylvania

COUNTRY: U.S.A.

ZIP: 19103

COMPUTER READABLE FORM: disk

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/266,408

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/089,986

FILING DATE: 08-JUL-1993

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 306 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-266-408-2

Query Match 44.0%; Score 37; DB 2; Length 306;

Best Local Similarity 31.2%; Pred. No. 68;

Matches 5; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

QY 1 LATRLMKFKAEIRDF 16

er

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 11:00:39 ; Search time 34.94 Seconds  
(without alignments)  
37.063 Million cell updates/sec

Title: US-09-763-397A-23  
Perfect score: 84  
Sequence: 1 LATRLMKKFAEIRDF 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	84	100.0	782	2 S27833	rhothy-associated
2	43	51.2	227	2 B82902	tRNA pseudouridine
3	43	51.2	353	2 T05552	SRG1 protein-relat
4	42	50.0	141	2 A32124	photosystem I chai
5	42	50.0	346	2 S73894	tryptophan--tRNA l
6	42	50.0	398	2 F85056	hypothetical prote
7	42	50.0	456	2 B71135	hypothetical prote
8	42	50.0	496	2 T52255	cytochrome P450 [i
9	42	50.0	743	2 C86168	hypothetical prote
10	41	48.8	319	2 H84808	probable annexin l
11	41	48.8	503	2 T52172	probable cytochrom
12	41	48.8	619	2 D69318	threonyl-tRNA synt
13	41	48.8	1530	1 S01393	DNA-directed RNA p
14	41	48.8	1625	2 T02921	acetyl-CoA carboxy
15	41	48.8	1685	2 T02750	acetyl-CoA carboxy
16	41	48.8	2325	2 T02235	acetyl-CoA carboxy
17	40	47.6	302	2 H83478	probable binding p
18	40	47.6	346	2 G86305	SRG1 homolog [impo
19	40	47.6	356	2 T05551	SRG1 protein-relat
20	40	47.6	527	2 H85230	hypothetical prote
21	40	47.6	535	2 T03246	cytochrome P450 -
22	40	47.6	535	2 T03260	cytochrome P450 -
23	40	47.6	568	2 S51275	DNA polymerase - p
24	40	47.6	774	1 VPXRWT	outer layer protei
25	40	47.6	775	1 VPXRW7	outer layer protei
26	40	47.6	776	1 VPXRW9	outer layer protei
27	40	47.6	1206	2 D84542	probable chloropia
28	39	46.4	268	2 F71712	prolipoprotein dia
29	39	46.4	344	2 C70388	hypothetical prote

SRG1 protein - Ara  
glutamate dehydrog  
2-dehydro-3-deoxyp  
nitrogen regulatio  
outer capsid prote  
HsdRI protein - My  
cytotoxin-associat  
genome polyprotein  
conserved hypoteth  
hypothetical prote  
translation elonga  
hypothetical prote  
cystathionine gamm  
2-dehydro-3-deoxyp  
probable DNA-direc

30	39	46.4	358	2 S44261
31	39	46.4	429	2 G84176
32	39	46.4	511	2 S40411
33	39	46.4	601	2 G71666
34	39	46.4	776	2 A48480
35	39	46.4	986	2 S49394
36	39	46.4	1215	2 B48281
37	39	46.4	2105	1 A44059
38	38	45.2	230	2 A83362
39	38	45.2	235	2 E86127
40	38	45.2	337	2 T06720
41	38	45.2	397	2 JH0416
42	38	45.2	398	2 F64456
43	38	45.2	509	2 A35016
44	38	45.2	541	2 S40412
45	38	45.2	547	2 S41618

ALIGNMENTS

RESULT 1  
S27833  
rhothy-associated protein 1 precursor - malaria parasite (Plasmodium falciparum)  
N:Alternate names: protective antigen  
C:Species: Plasmodium falciparum  
C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 09-Jun-2000  
C:Accession: A45514; S27833  
R:Ridley, R.G.; Takacs, B.; Lahm, H.W.; Delves, C.J.; Goman, M.; Certa, U.; Matile, H.  
Mol. Biochem. Parasitol. 41, 125-134, 1990  
A:Title: Characterisation and sequence of a protective rhothy antigen from Plasmodiu  
A:Reference number: A45514; MUID:90348711  
A:Accession: A45514  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-782 <RI2>  
A:Cross-references: GB:M32853; NID:g160656; PID:g160657  
C:Superfamily: Plasmodium falciparum rhothy-associated protein 1

Query Match 100.0%; Score 84; DB 2; Length 782;  
Best Local Similarity 100.0%; Pred. No. 8e-06;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LATRLMKKFAEIRDF 17  
DB 736 LATRLMKKFAEIRDF 752  
|||||  
|||||

RESULT 2  
B82902  
tRNA pseudouridine 55 synthase U0354 [imported] - Ureaplasma urealyticum  
C:Species: Ureaplasma urealyticum  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C:Accession: B82902  
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.  
submitted to GenBank, February 2000  
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a  
A:Reference number: A82870  
A:Accession: B82902  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-227 <GLA>  
A:Cross-references: GB:AE002132; GB:AF222894; NID:g6899327; PIDN:AAF30763.1; GSPDB:GN  
A:Experimental source: serovar 3; biovar 1  
C:Genetics:  
A:Gene: trbB; U0354  
A:Genetic code: SGC3

Query Match 51.2%; Score 43; DB 2; Length 227;  
Best Local Similarity 56.2%; Pred. No. 11;  
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Query Match 51.2%; Score 43; DB 2; Length 353;  
Best Local Similarity 58.3%; Pred. No. 18;  
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 6 MKKFKAEIRDF 17  
: |||: |||  
Db 92 LDKFKSDIQDF 103

RESULT 4  
A32124  
photosystem I chain II - Synechocystis sp. (strain PCC 6803)  
N:Alternate names: photosystem I reaction center chain II  
C:Species: Synechocystis sp.  
A:Variety: PCC 6803  
C:Date: 07-Jun-1990 #sequence\_revision 07-Jun-1990 #text\_change 20-Jun-2000  
C:Accession: A32124; S74536; S67980  
R:Reilly, P.; Hulmes, J.D.; Pan, Y.C.E.; Nelson, N.  
J. Biol. Chem. 263, 17658-17662, 1988  
A:Title: Molecular cloning and sequencing of the psad gene encoding subunit II of photosystem I  
A:Reference number: A92692; MUID:89034300  
A:Accession: A32124  
A:Molecule type: DNA  
A:Residues: 1-141 <REI>  
A:Cross-references: GB:J04195; NID:g154480; PIDN:AAA88625.1; PID:g154481  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, K. DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp.  
S:Reference number: S74322; MUID:97061201  
A:Accession: S74536  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-141 <KAN>  
A:Cross-references: EMBL:D90899; GB:AB001339; NID:g1651650; PIDN:BAA16688.1; PID:g165176  
R:Barbato, R.; Polverino de Laureto, P.; Rigoni, F.; de Martini, E.; Giacometti, G.M. Eur. J. Biochem. 234, 459-465, 1995  
A:Title: Pigment-protein complexes from the photosynthetic membrane of the cyanobacterium Synechocystis sp. PCC 6803  
A:Reference number: S67978; MUID:96128174  
A:Accession: S67980  
A:Molecule type: protein  
A:Residues: 2, 'S', 4-9, 'L', 11-13 <BAR>  
C:Genetics:  
A:Gene: psad

Query Match 50.0%; Score 42; DB 2; Length 141;  
Best Local Similarity 50.0%; Pred. No. 11;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 LATRLMKKFKAEIRDF 16  
: |||: |||: |||  
Db 68 LGTQLTKFKPKIQDY 83

RESULT 5  
S73894  
trypthophan--trna ligase (EC 6.1.1.2) trps - Mycoplasma pneumoniae (strain ATCC 29342)  
N:Alternate names: hypothetical protein A65\_orf346; tryptophanyl-trna synthetase trps  
C:Species: Mycoplasma pneumoniae  
A:Variety: ATCC 29342  
C:Date: 27-Feb-1997 #sequence\_revision 25-Apr-1997 #text\_change 07-Dec-1999  
C:Accession: S73894  
R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R. Nucleic Acids Res. 24, 4420-4449, 1996  
A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae  
A:Reference number: S73327; MUID:97105885  
A:Accession: S73894  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-346 <HIM>  
A:Cross-references: EMBL:AE000056; GB:U00089; NID:g1674263; PIDN:AAB96216.1; PID:g1674263  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996  
C:Genetics:  
A:Gene: trps  
A:Genetic code: SGC3  
C:Superfamily: tryptophan--trna ligase  
C:Keywords: aminoacyl-trna synthetase; ligase; protein biosynthesis

Query Match 50.0%; Score 42; DB 2; Length 346;  
Best Local Similarity 57.1%; Pred. No. 25;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 LATRLMKKFKAEIR 14  
: |||: |||: |||  
Db 165 LAQRIQKKFKLKL 178

RESULT 6  
F85056  
hypothetical protein AT4g04480 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: F85056  
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Harbor Laboratory, New York, 1999  
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A:Reference number: A85001; MUID:20083488  
A:Accession: F85056  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-398 <STO>  
A:Cross-references: GB:NC\_001268; NID:g7267205; PIDN:CAB77916.1; GSPDB:GN00140  
C:Genetics:  
A:Gene: AT4g04480  
A:Map position: 4

Query Match 50.0%; Score 42; DB 2; Length 398;  
Best Local Similarity 60.0%; Pred. No. 29;  
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 ATRLMKKFKAEIRDF 16  
: |||: |||: |||  
Db 165 ATRLFKKLRLVEIMF 179

C:Superfamily: photosystem I chain II

Query Match 50.0%; Score 42; DB 2; Length 141;  
Best Local Similarity 50.0%; Pred. No. 11;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 LATRLMKKFKAEIRDF 16  
: |||: |||: |||  
Db 68 LGTQLTKFKPKIQDY 83

RESULT 5  
S73894  
trypthophan--trna ligase (EC 6.1.1.2) trps - Mycoplasma pneumoniae (strain ATCC 29342)  
N:Alternate names: hypothetical protein A65\_orf346; tryptophanyl-trna synthetase trps  
C:Species: Mycoplasma pneumoniae  
A:Variety: ATCC 29342  
C:Date: 27-Feb-1997 #sequence\_revision 25-Apr-1997 #text\_change 07-Dec-1999  
C:Accession: S73894  
R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R. Nucleic Acids Res. 24, 4420-4449, 1996  
A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae  
A:Reference number: S73327; MUID:97105885  
A:Accession: S73894  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-346 <HIM>  
A:Cross-references: EMBL:AE000056; GB:U00089; NID:g1674263; PIDN:AAB96216.1; PID:g1674263  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996  
C:Genetics:  
A:Gene: trps  
A:Genetic code: SGC3  
C:Superfamily: tryptophan--trna ligase  
C:Keywords: aminoacyl-trna synthetase; ligase; protein biosynthesis

Query Match 50.0%; Score 42; DB 2; Length 346;  
Best Local Similarity 57.1%; Pred. No. 25;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 LATRLMKKFKAEIR 14  
: |||: |||: |||  
Db 165 LAQRIQKKFKLKL 178

RESULT 6  
F85056  
hypothetical protein AT4g04480 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: F85056  
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Harbor Laboratory, New York, 1999  
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A:Reference number: A85001; MUID:20083488  
A:Accession: F85056  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-398 <STO>  
A:Cross-references: GB:NC\_001268; NID:g7267205; PIDN:CAB77916.1; GSPDB:GN00140  
C:Genetics:  
A:Gene: AT4g04480  
A:Map position: 4

Query Match 50.0%; Score 42; DB 2; Length 398;  
Best Local Similarity 60.0%; Pred. No. 29;  
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 ATRLMKKFKAEIRDF 16  
: |||: |||: |||  
Db 165 ATRLFKKLRLVEIMF 179



```
RESULT 7
B71135
hypothetical protein PH0846 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C:Accession: B71135
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekinaka, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, M.; Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaeon
A:Reference number: A71000; MUID:98344137
A:Accession: B71135
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-456 <KAW>
A:Cross-references: GB:AF000003; NID:93236130; PIDN:BAA29940.1; PID:93257257
A:Experimental source: strain OT3
A:Note: This accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH0846
C:Superfamily: Methanobacterium thermoautotrophicum conserved hypothetical protein MTH19

Query Match 50.0%; Score 42; DB 2; Length 456;
Best Local Similarity 53.3%; Pred. No. 33;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TRLMKFKAEIRDF 17
|||:|:|:|
Db 35 TRLMKFKAEIRDF 49

RESULT 8
T52255
cytochrome P450 [imported] - Thlaspi arvense
C:Species: Thlaspi arvense
C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 16-Feb-2001
C:Accession: T52255
R:Udvardi, M.K.; Metzger, J.D.; Krishnapillai, V.; Peacock, W.J.; Dennis, E.S.
Plant Physiol. 104, 755-756, 1994
A:Title: Cloning and nucleotide sequence of a full-length cDNA from Thlaspi arvense that encodes cytochrome P450
A:Reference number: Z26005
A:Accession: T52255
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-496 <UDV>
A:Cross-references: EMBL:L24438; PIDN:AAAL19701.1
C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C:Keywords: heme; iron; metalloprotein
F:294-458/Domain: cytochrome P450 homology <P45>
F:436/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 50.0%; Score 42; DB 2; Length 496;
Best Local Similarity 66.7%; Pred. No. 35;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 4 RLMKKFKAEIRD 15
|||:|:|:|
Db 318 RVMKKLQAEIRE 329

RESULT 9
C86168
hypothetical protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C86168
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Chiu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizlar, L.

Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maithi, R.; Marzia Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Talloker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: C86168
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-743 <STO>
A:Cross-references: GB:AE005172; NID:g4204309; PIDN:AAD10690.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 50.0%; Score 42; DB 2; Length 743;
Best Local Similarity 61.5%; Pred. No. 52;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 5 LMKKFKAEIRDF 17
|||:|:|:|
Db 285 LMKKFKAEIRDF 297

RESULT 10
H84808
probable annexin [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: H84808
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: H84808
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-319 <STO>
A:Cross-references: GB:AE002093; NID:g3785997; PIDN:AAC67343.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g38750
A:Map position: 2

Query Match 48.8%; Score 41; DB 2; Length 319;
Best Local Similarity 50.0%; Pred. No. 34;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 1 LATRLMKFKAEIRDF 16
|||:|:|:|
Db 295 LAQRIQEKIKGNRDF 310

RESULT 11
T52172
probable cytochrome P450 At2g24180 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 16-Feb-2001
C:Accession: T52172; E84633
R:Mizutani, M.; Ward, E.; Ohta, D.
Plant Mol. Biol. 37, 39-52, 1998
A:Title: Cytochrome P450 superfamily in Arabidopsis thaliana: isolation of cDNAs, dif
A:Reference number: Z14382; MUID:98281573
A:Accession: T52172
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-503 <MI>
A:Cross-references: EMBL:D78604; PIDN:BAA28536.1
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
```

```
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maithi, R.; Marzia Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Talloker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: C86168
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-743 <STO>
A:Cross-references: GB:AE005172; NID:g4204309; PIDN:AAD10690.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 50.0%; Score 42; DB 2; Length 743;
Best Local Similarity 61.5%; Pred. No. 52;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 5 LMKKFKAEIRDF 17
|||:|:|:|
Db 285 LMKKFKAEIRDF 297

RESULT 10
H84808
probable annexin [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: H84808
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: H84808
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-319 <STO>
A:Cross-references: GB:AE002093; NID:g3785997; PIDN:AAC67343.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g38750
A:Map position: 2

Query Match 48.8%; Score 41; DB 2; Length 319;
Best Local Similarity 50.0%; Pred. No. 34;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 1 LATRLMKFKAEIRDF 16
|||:|:|:|
Db 295 LAQRIQEKIKGNRDF 310

RESULT 11
T52172
probable cytochrome P450 At2g24180 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 16-Feb-2001
C:Accession: T52172; E84633
R:Mizutani, M.; Ward, E.; Ohta, D.
Plant Mol. Biol. 37, 39-52, 1998
A:Title: Cytochrome P450 superfamily in Arabidopsis thaliana: isolation of cDNAs, dif
A:Reference number: Z14382; MUID:98281573
A:Accession: T52172
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-503 <MI>
A:Cross-references: EMBL:D78604; PIDN:BAA28536.1
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
```

M.: Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
euss, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487  
A:Accession: E84633  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-503 <STO>  
A:Cross-references: GB:AE002093; NID:g4115378; PIDN:RAD03379.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g24180; CYP71B6  
A:Map position: 2  
C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology  
C:Keywords: heme; iron; metalloprotein  
F:304-468/Domain: cytochrome P450 homology <P45>  
F:446/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 48.8%; Score 41; DB 2; Length 503;  
Best Local Similarity 66.7%; Pred. No. 52;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
Db 328 RVMKVQAEIRE 339

QY 4 RLKMKKFAEIRD 15  
|:|:|:|:|:|:  
Db 328 RVMKVQAEIRE 339

RESULT 12  
D69318  
C:Species: Archaeoglobus fulgidus  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 08-Oct-1999  
C:Accession: D69318  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.  
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kalne, B.P.; Sykes, S.  
Smith, H.O.; Woese, C.R.; Venter, J.C.  
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae  
A:Reference number: A69250; MUID:98049343  
A:Accession: D69318  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-619 <KLE>  
A:Cross-references: GB:AE001066; GB:AE000782; NID:g2689389; PIDN:AAB90685.1; PID:g265007  
C:Superfamily: Methanococcus jannaschii threonine--trna ligase

Query Match 48.8%; Score 41; DB 2; Length 619;  
Best Local Similarity 46.7%; Pred. No. 64;  
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;  
Db 88 AVKLKQLEALSDF 102

QY 2 ATRLMKKFAEIRDF 16  
|:|:|:|:|:|:  
Db 88 AVKLKQLEALSDF 102

RESULT 13  
S01393  
DNA-directed RNA polymerase (EC 2.7.7.6) III 170K chain - Trypanosoma brucei  
C:Species: Trypanosoma brucei  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: S01393  
R:Koeck, J.; Evers, R.; Cornelissen, A.W.C.A.  
Nucleic Acids Res. 16, 8753-8772, 1988  
A:Title: Structure and sequence of the gene for the largest subunit of trypanosomal RNA  
A:Reference number: S01393; MUID:89016560  
A:Accession: S01393  
A:Molecule type: DNA  
A:Residues: 1-1530 <KOE>  
A:Cross-references: EMBL:X12494; NID:g10523; PIDN:CAA31014.1; PID:g10524

C:Superfamily: yeast DNA-directed RNA polymerase III 160K chain  
C:Keywords: DNA binding; nucleotidyltransferase; nucleus; transcription; zinc finger

Query Match 48.8%; Score 41; DB 1; Length 1530;  
Best Local Similarity 63.6%; Pred. No. 1.5e+02;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
QY 7 KFKFAEIRDFF 17  
|:|:|:|:|:|:  
Db 1035 RKFKDDIQDFF 1045

RESULT 14  
T02921  
acetyl-CoA carboxylase (EC 6.4.1.2) (clone A3) - maize (fragment)  
C:Species: Zea mays (maize)  
C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 21-Jul-2000  
C:Accession: T02921  
R:Ashton, A.R.; Jenkins, C.L.D.; Whitfield, P.R.  
Plant Mol. Biol. 24, 35-49, 1994  
A:Title: Molecular cloning of two different cDNAs for maize acetyl CoA carboxylase.  
A:Reference number: S42659; MUID:94154242  
A:Accession: T02921  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-1625 <ASH>  
A:Cross-references: EMBL:Z24449; NID:g1272684; PIDN:CAA80822.1; PID:g1272685  
A:Experimental source: strain B73; leaf  
C:Function:  
A:Description: catalyzes the ATP-dependent carboxylation of acetyl-CoA to malonyl-CoA  
C:Superfamily: human acetyl-CoA carboxylase; biotin carboxylase homology; lipoyl/biot  
C:Keywords: ligase  
F:68-140/Domain: lipoyl/biotin-binding homology <LPB>

Query Match 48.8%; Score 41; DB 2; Length 1625;  
Best Local Similarity 53.3%; Pred. No. 1.6e+02;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 1 LATRLMKKFAEIRD 15  
|:|:|:|:|:|:  
Db 220 LATRLPRNLKSELED 234

RESULT 15  
T02750  
acetyl-CoA carboxylase (EC 6.4.1.2) - maize (fragment)  
C:Species: Zea mays (maize)  
C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 21-Jan-2000  
C:Accession: T02750  
R:Ashton, A.R.  
submitted to the EMBL Data Library, April 1996  
A:Reference number: Z14716  
A:Accession: T02750  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-1685 <ASH>  
A:Cross-references: EMBL:U58598; NID:g1335965; PIDN:AAB01188.1; PID:g1335966  
C:Function:  
A:Description: catalyzes the ATP-dependent carboxylation of acetyl-CoA to malonyl-CoA  
A:Pathway: fatty acid biosynthesis  
A:Note: biotin cofactor  
C:Superfamily: human acetyl-CoA carboxylase; biotin carboxylase homology; lipoyl/biot  
C:Keywords: ligase  
F:127-199/Domain: lipoyl/biotin-binding homology <LPB>

Query Match 48.8%; Score 41; DB 2; Length 1685;  
Best Local Similarity 53.3%; Pred. No. 1.7e+02;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 1 LATRLMKKFAEIRD 15

Db 279 LATRLPRNLKSELED 293

Search completed: January 29, 2002, 11:00:40  
Job time: 2030 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 11:16:19 ; Search time 20.36 Seconds  
(without alignments)  
30.614 Million cell updates/sec

Title: US-09-763-397A-23

Perfect score: 84

Sequence: 1 LATRLMKFKAEIRDF 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	43	51.2	227	1	TRUB_UREPA	O9pqd7 ureaplasma
2	43	51.2	501	1	ODPX_HUMAN	O00330 homo sapien
3	42	50.0	140	1	PSAD_SYNY3	P19569 synchocyst
4	42	50.0	346	1	SYW_MYCPN	P75510 mycoplasma
5	42	50.0	496	1	C7B1_THLAR	P49264 thlaspi arv
6	41	48.8	503	1	C726_ATH	O65787 arabidopsis
7	41	48.8	619	1	SYT_ARCFU	O23703 archaeoglob
8	41	48.8	1530	1	RPC1_TRYBB	P08968 trypanosoma
9	40	47.6	194	1	YMP9_CAEEL	Q10952 caenorhabdi
10	40	47.6	500	1	C72Q_ATH	O91t10 arabidopsis
11	40	47.6	535	1	C7C3_MAI2E	P93703 zea mays (m
12	40	47.6	774	1	VP4_ROTHT	P11200 human rotav
13	40	47.6	775	1	VP4_ROTTH	P11197 human rotav
14	40	47.6	776	1	VP4_ROTTH	P11199 human rotav
15	39	46.4	268	1	LST_RICPR	O9ze99 rickettsia
16	39	46.4	344	1	YA22_AQUAE	O67134 aquifex aeo
17	39	46.4	429	1	DHE4_HALN1	Q9bsm4 halobacteri
18	39	46.4	496	1	C72B_ATH	P58049 arabidopsis
19	39	46.4	496	1	C72D_ATH	P58050 arabidopsis
20	39	46.4	496	1	C72E_ATH	P58051 arabidopsis
21	39	46.4	511	1	AROF_LYCES	P37215 lycopersico
22	39	46.4	1147	1	CGA1_HELPY	P80200 helicobacte
23	39	46.4	2105	1	POLR_ASGVP	P36309 apple stem
24	38	45.2	397	1	EFTU_MYCHO	P22679 mycoplasma
25	38	45.2	502	1	C72G_ATH	Q91tm7 arabidopsis
26	38	45.2	538	1	AROF_SOLTU	P21357 solanum tub
27	38	45.2	541	1	AROF_LYCES	P27216 lycopersico
28	38	45.2	542	1	AROF_TOBAC	P27608 nicotiana t
29	38	45.2	775	1	VP4_ROTGP	P23045 porcine rot
30	38	45.2	776	1	VP42_ROTSL	P12976 simian 11 r
31	38	45.2	776	1	VP42_ROTBS	P36305 bovine rota
32	38	45.2	776	1	VP4_ROT9	O06894 canine rota
33	38	45.2	776	1	VP4_ROT6	O07416 feline rota

34	38	45.2	776	1	VP4_ROT6	P26451 human rotav
35	38	45.2	776	1	VP4_ROT6	O06895 human rotav
36	38	45.2	776	1	VP4_ROTSS	P17464 simian 11 r
37	37.5	44.6	1323	1	LT23_CAEEL	P24348 caenorhabdi
38	37	44.0	187	1	RL5_MYCTU	P95064 mycobacteri
39	37	44.0	189	1	YNO0_YEAST	P53910 saccharomyc
40	37	44.0	244	1	RHOE_HUMAN	P52199 homo sapien
41	37	44.0	244	1	RHOE_PIG	O77683 sus scrofa
42	37	44.0	263	1	TRUA_PYRAB	Q9uz23 pyrococcus
43	37	44.0	322	1	ACC1_ORYSA	O40634 oryza sativ
44	37	44.0	329	1	NUC1_YEAST	P08466 saccharomyc
45	37	44.0	347	1	SYW_MYCGE	P47372 mycoplasma

ALIGNMENTS

RESULT 1  
TRUB\_UREPA  
ID TRUB\_UREPA STANDARD; PRT; 227 AA.  
AC Q9POD7: 20-AUG-2001 (Rel. 40, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE TRNA PSEUDOURIDINE SYNTHASE B (EC 4.2.1.70) (TRNA PSEUDOURIDINE 5S  
DE SYNTHASE) (PSI55 SYNTHASE) (PSEUDOURIDYLATE SYNTHASE) (URACIL  
DE HYDROLYASE).  
GN TRUB OR UU354.  
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
OC Mycoplasmataceae; Ureaplasma.  
OX NCBI\_TaxID=134821;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SEROVAR 3;  
RX MEDLINE=20500219; PubMed=11048724;  
RA Glass J.I., Lefkowitz E.J., Glass S.J., Helner C.R., Chen E.Y.,  
RA Casseil G.H.;  
RT "The complete sequence of the mucosal pathogen Ureaplasma  
urealyticum".  
RT Nature 407:757-762(2000).  
CC -!- FUNCTION: FORMATION OF PSEUDOURIDINE AT POSITION 55 IN THE  
CC PSI GC LOOP OF TRANSFER RNAS (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: URACIL + D-RIBOSE 5-PHOSPHATE = PSEUDOURIDINE  
CC 5'-PHOSPHATE + H(2)O.  
CC -!- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL: AE002132; AAF30763.1; -  
CC InterPro: IPR002501; TruB\_N.  
DR Pfam: PF01509; TruB\_N; 1.  
KW Lyase; tRNA processing; Complete proteome.  
SQ SEQUENCE 227 AA; 25765 MW; FAICADC357D6AD36 CRC64;

Query Match 51.2%; Score 43; DB 1; Length 227;  
Best Local Similarity 56.2%; Pred No. 3.7;  
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 1 LATRLMKFKAEIRDF 16  
: ||: ||: ||  
Db 143 IAPRLVKFKYQIMDF 158

RESULT 2  
ODPX\_HUMAN

ID ODPX\_HUMAN STANDARD; PRT; 501 AA.  
AC 000330; Q99783; 060221;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE PYRUVATE DEHYDROGENASE PROTEIN X COMPONENT, MITOCHONDRIAL PRECURSOR  
DE (DIHYDROLIPOAMIDE DEHYDROGENASE-BINDING PROTEIN OF PYRUVATE  
DE DEHYDROGENASE COMPLEX) (E3-BINDING PROTEIN) (E3BP) (PROX).  
GN PDX1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=973930399; PubMed=9242632;  
RX Harris R.A., Bowker-Kinley M.M., Wu P., Jeng J., Popov K.M.;  
RA "Dihydrolipoamide dehydrogenase-binding protein of the human pyruvate  
RT dehydrogenase complex. DNA-derived amino acid sequence, expression,  
RT and reconstitution of the pyruvate dehydrogenase complex.";  
RL J. Biol. Chem. 272:19746-19751(1997).  
[2]  
RN SEQUENCE FROM N.A.  
RP TISSUE=Liver;  
RX MEDLINE=98153629; PubMed=9399911;  
RA Aral B., Benelli C., Ait-Ghezala G., Amessou M., Fouque F.,  
RA Maunoury C., Creau N., Kamoun P., Marsac C.;  
RT "Mutations in PDX1, the human lipoyl-containing component X of the  
RT pyruvate dehydrogenase-complex gene on chromosome 11p1, in congenital  
RT lactic acidosis.";  
RL Am. J. Hum. Genet. 61:1318-1326(1997).  
[3]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=98133932; PubMed=9467010;  
RX Ling M., McEachern G., Seyda A., Mackay N., Scherer S.W.,  
RA Bratinova S., Beatty B., Giovannucci-Uzielli M.L., Robinson B.H.; X  
RT "Detection of a homozygous four base pair deletion in the protein  
RT gene in a case of pyruvate dehydrogenase complex deficiency.";  
RL Hum. Mol. Genet. 7:501-505(1998).  
[4]  
RN SEQUENCE FROM N.A.  
RP Moore M.;  
RA Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
[5]  
RN SEQUENCE OF 127-501 FROM N.A.  
RP TISSUE=Brain;  
RX MEDLINE=97264341; PubMed=9110174;  
RA Yu W., Andersson B., Worley K.C., Muzny D.M., Ding Y., Liu W.,  
RA Ricafrente J.Y., Wentland M.A., Lennon G., Gibbs R.A.;  
RT "Large-scale concatenation cDNA sequencing.";  
RL Genome Res. 7:353-358(1997).  
CC -!- FUNCTION: REQUIRED FOR ANCHORING DIHYDROLIPOAMIDE DEHYDROGENASE  
CC (E3) TO THE DIHYDROLIPOAMIDE TRANSACETYLASE (E2) CORE OF THE  
CC PYRUVATE DEHYDROGENASE COMPLEXES OF EUKARYOTES. THIS SPECIFIC  
CC BINDING IS ESSENTIAL FOR A FUNCTIONAL PDH COMPLEX.  
CC -!- SUBUNIT: EUKARYOTIC PYRUVATE DEHYDROGENASE COMPLEXES ARE ORGANIZED  
CC ABOUT A CORE CONSISTING OF THE OLIGOMERIC DIHYDROLIPOAMIDE ACETYL-  
CC TRANSFERASE, AROUND WHICH ARE ARRANGED MULTIPLE COPIES OF PYRUVATE  
CC DEHYDROGENASE, DIHYDROLIPOAMIDE DEHYDROGENASE AND PROTEIN X BOUND  
CC BY NONCOVALENT BONDS.  
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.  
CC -!- DISEASE: DEFECTS IN PDX1 ARE A CAUSE OF LACTICACIDEMIA.  
CC -!- SIMILARITY: BELONGS TO THE 2-OXOACID DEHYDROGENASE FAMILY.  
CC -!- SIMILARITY: CONTAINS 1 LIPOYL-BINDING DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

DR EMBL; AF001437; AAB66315.1; -.  
DR EMBL; Y13145; CAA73606.1; -.  
DR EMBL; U82328; AAC39661.1; -.  
DR EMBL; AL138810; CAC12641.1; -.  
DR EMBL; U79296; AAB50223.1; -.  
DR MIM; 245349; -.  
DR InterPro; IPR001078; 2oxoacid\_dh.  
DR InterPro; IPR000089; Biotin\_lipoyl.  
DR InterPro; IPR003016; Lipoyl.  
DR InterPro; IPR002965; P-rich\_extensn.  
DR Pfam; PF00198; 2-oxoacid\_dh; 1.  
DR Pfam; PF00364; biotin\_lipoyl; 1.  
DR PRINTS; PR01217; PRICHEXTENSN.  
DR PRODOM; PD001115; 2oxoacid\_dh; 1.  
DR PROSITE; PS00189; LIPOYL; 1.  
KW Transit peptide; Mitochondrion; Lipoyl.  
FT TRANSIT 1 53 MITOCHONDRION (BY SIMILARITY).  
FT CHAIN 54 501 PYRUVATE DEHYDROGENASE PROTEIN X  
FT COMPONENT.  
FT BINDING 97 97 LIPOYL (BY SIMILARITY).  
FT CONFLICT 23 23 R -> C (IN REF. 1).  
FT CONFLICT 41 41 A -> R (IN REF. 3).  
FT CONFLICT 251 251 A -> S (IN REF. 1 AND 2).  
SQ SEQUENCE 501 AA; 54122 MW; 9CF0C1DAE9E12EF9 CRC64;

Query Match 51.2%; Score 43; DB 1; Length 501;  
Best Local Similarity 53.3%; Pred. No. 8.4;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 LATRLMKKFAEIRD 15  
| | | | : | | | : :  
- Db 482 LATRLKSKFANLEN 496

RESULT 3  
PSAD\_SYNV3  
ID PSAD\_SYNV3 STANDARD; PRT; 140 AA.  
AC P19569;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE PHOTOSYSTEM I REACTION CENTRE SUBUNIT II (PHOTOSYSTEM I 16 KDA  
DE POLYPEPTIDE) (PSI-D).  
DE PSAD OR SLR0737.  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
OX NCBI\_TaxID=1148;  
[1]  
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=89034300; PubMed=3141423;  
RA Reilly P., Hulmes J.D., Pan Y.-C.E., Nelson N.;  
RT "Molecular cloning and sequencing of the psad gene encoding subunit  
RT II of photosystem I from the cyanobacterium, Synechocystis sp. PCC  
RT 6803.";  
RL J. Biol. Chem. 263:17658-17662(1988).  
[2]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=97061201; PubMed=8905231;  
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
RA Miyajima N., Hiroseawa M., Sugiyama M., Sasamoto S., Kimura T.,  
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,  
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,  
RA Yamada M., Yasuda M., Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the  
RT entire genome and assignment of potential protein-coding regions.";  
RL DNA Res. 3:109-136(1996).  
CC -!- FUNCTION: PSAD CAN FORM COMPLEXES WITH FERREDOXIN AND FERREDOXIN-  
CC OXIDOREDUCTASE IN PHOTOSYSTEM I (PS I) REACTION CENTRE.  
CC -!- SUBUNIT: PS I FROM SYNECHOCYSTIS SP. CONTAINS SEVEN DIFFERENT  
CC POLYPEPTIDES. THE POLYPEPTIDE CORRESPONDING TO SUBUNIT II FROM  
CC HIGHER PLANTS IS MISSING.

```
CC -!- SIMILARITY: BELONGS TO THE PSAD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J04195; AAA88625.1; .
DR EMBL; D90899; BAA16688.1; .
DR PIR; A32124; A32124.
DR InterPro; IPR003685; Psad.
DR Pfam; PF02531; Psad; 1.
KW Photosynthesis; Photosystem I; Complete proteome.
FT INIT_MET 0
SQ SEQUENCE 140 AA; 15513 MW; 263C35D5B47B9F8E CRC64;

Query Match 50.0%; Score 42; DB 1; Length 140;
Best Local Similarity 50.0%; Pred. No. 3.3;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 LATRLMKKFKAEIRDF 16
| | | | | | | |
Db 67 LCTQLTRFKPKIQDY 82

RESULT 4
SYW_MYCPN STANDARD; PRT; 346 AA.
AC P75510;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRYPTOPHANYL-TRNA SYNTHETASE (EC 6.1.1.2) (TRYPTOPHAN--TRNA LIGASE)
DE (TRPRS).
GN TRPS OR MPN265 OR MP568.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 29342 / M129;
RX MEDLINE=97103685; PubMed=8948633;
RA Himmelreich R., Hilbert H., Piagens H., Pirkel E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae."
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -!- CATALYTIC ACTIVITY: ATP + L-TRYPTOPHAN + TRNA(TRP) = AMP +
CC PYROPHOSPHATE + L-TRYPTOPHANYL-TRNA(TRP).
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000056; AAB96216.1; .
DR InterPro; IPR002305; tRNA-synt_1b.
DR InterPro; IPR001412; tRNA-synt_1.
DR Pfam; PF00579; tRNA-synt_1b; 1.
DR PRINTS; PR01039; TRNASYNTHTRP.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; FALSE_NEG.
```

```
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 11 19 "HIGH" REGION.
FT SITE 200 204 "RMSKS" REGION.
FT BINDING 203 203 ATP (BY SIMILARITY).
SQ SEQUENCE 346 AA; 39079 MW; 56CF4FFCCB9C2F95 CRC64;

Query Match 50.0%; Score 42; DB 1; Length 346;
Best Local Similarity 57.1%; Pred. No. 8.4;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 LATRLMKKFKAEIR 14
| | | | | | | |
Db 165 LAQRIQKKFKLKL 178

RESULT 5
C7BL_THLAR STANDARD; PRT; 496 AA.
ID C7BL_THLAR
AC P49264;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME P450 71B1 (EC 1.14.-.-) (CYPLXXIB1).
DE CYP71B1.
GN CYP71B1.
OS Thlaspi arvense (Field pennycress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Thlaspi.
OX NCBI_TaxID=13288;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Shoot apex;
RX MEDLINE=94345014; PubMed=8066138;
RA Udvardi M.K., Metzger J.D., Krishnapillai V., Peacock W.J.,
RA Dennis E.S.;
RT "Cloning and sequencing of a full-length cDNA from Thlaspi arvense L.
RT that encodes a cytochrome P-450."
RL Plant Physiol. 105:755-756(1994).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L24438; AAA19701.1; .
DR InterPro; IPR001128; Cyt_P450.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR00463; EP4501.
DR PROSITE; PS00086; CYTOCHROME P450; 1.
KW Oxidoreductase; Monooxygenase; Heme.
FT BINDING 436 436 HEME (BY SIMILARITY).
SQ SEQUENCE 496 AA; 56387 MW; 59D3A6B34F668D08 CRC64;

Query Match 50.0%; Score 42; DB 1; Length 496;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 RLMKKFKAEIRD 15
| | | | | | | |
Db 318 RYMKKIQAEIRE 329

RESULT 6
C726_ARATH
ID C726_ARATH STANDARD; PRT; 503 AA.
AC O65787;
```

DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE CYTOCHROME P450 71B6 (EC 1.14.-.-).  
 GN CP71B6 OR AT2G24180 OR F27D4.9.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, COLUMBIA; TISSUE=Seedling;  
 RX MEDLINE=96281573; PubMed=9620263;  
 RA Mizutani M., Ward E., Ohta D.;  
 RA "Cytochrome P450 superfamily in Arabidopsis thaliana: isolation of  
 RT cDNAs, differential expression, and RFLP mapping of multiple  
 FT cytochromes P450.";  
 RL Plant Mol. Biol. 37:39-52(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, COLUMBIA;  
 RX MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,  
 RA Moffat K.S., Cronin L.A., Shen M., VanAken S.E., Umayam L.,  
 RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,  
 RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,  
 RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,  
 RA Venter J.C.;  
 RA "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 RT thaliana.";  
 RL Nature 402:761-768(1999).  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D78604; BAA28536.1; -.  
 DR EMBL; AC005967; AAD03379.1; -.  
 DR InterPro; IPR001128; Cyt\_P450.  
 DR Pfam; PF00067; p450; 1.  
 DR PRINTS; PS00463; EP4501.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
 KW Oxidoreductase; Monooxygenase; Transmembrane; Heme; Multigene family.  
 FT TRANSMEM 10 30 POTENTIAL.  
 FT BINDING 446 446 HEME (BY SIMILARITY).  
 FT SEQUENCE 503 AA; 57008 MW; 71AFAD5370AEDFF6 CRC64;  
 SQ  
 Query Match 48.8%; Score 41; DB 1; Length 503;  
 Best Local Similarity 66.7%; Pred. No. 18;  
 Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 QY 4 RLMKKFAEIRD 15  
 DB 328 RVMKVQAEIRE 339  
 RESULT 7  
 ID SYT\_ARCFU STANDARD; PRT; 619 AA.  
 AC O29703;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE THREONYL-TRNA SYNTHETASE (EC 6.1.1.3) (THREONINE--TRNA LIGASE)

(THRES).  
 DE THRS OR AF0548.  
 OS Archaeoglobus fulgidus.  
 OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;  
 OC Archaeoglobus.  
 OX NCBI\_TaxID=2234;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
 RX MEDLINE=98049343; PubMed=9389475;  
 RA Klenk H.-P., Clayton R.A., Twinn M., Hickey E.K., Peterson J.D.,  
 RA Ketchum K.A., Dodson R.J., Gwinn M., Graham D.E., Kyrpides N.C.,  
 RA Richardson D.L., Kervatage A.R., Lee N.H., Sutton G.G., Gill S.,  
 RA Fleischmann R.D., Quackenbush J., McKenney K., Adams M.D., Loftus B.,  
 RA Kirkness E.F., Dougherty B.A., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
 RA Peterson S., Reich C.I., Weidman J.F., McDonald L., Utterback T.,  
 RA Overbeek R., Gocayne J.D., Weidman J.F., McKenney K., Kaine B.P., Sykes S.M.,  
 RA Cotton M.D., Spriggs T., Artiach P., Bowman C., Fujii C., Garland S.A.,  
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fraser C.M., Smith H.O., Woese C.R.,  
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 RA Venter J.C.;  
 RA "The complete genome sequence of the hyperthermophilic, sulphate-  
 RT reducing archaeon Archaeoglobus fulgidus.";  
 RL Nature 390:364-370(1997).  
 CC -1- CATALYTIC ACTIVITY: ATP + L-THREONINE + TRNA(THR) = AMP +  
 CC PYROPHOSPHATE + L-THREONYL-TRNA(THR).  
 CC -1- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE001066; AAB90685.1; -.  
 DR TIGR; AF0548; -.  
 DR InterPro; IPR002106; AA-trna\_ligase\_II.  
 DR InterPro; IPR002314; trna-synt\_2b.  
 DR InterPro; IPR002320; trna-synt\_thr.  
 DR Pfam; PF00587; trna-synt\_2b; 1.  
 DR PRINTS; PR01047; TRNASYNTHTR.  
 DR PROSITE; PS00179; AA-TRNA-LIGASE\_II\_1; 1.  
 DR PROSITE; PS00339; AA-TRNA-LIGASE\_II\_2; FALSE-NEG.  
 KW Aminoacyl-TRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 KW Metal-binding; Zinc; Complete proteome.  
 FT DOMAIN 196 495 CATALYTIC.  
 FT METAL 288 288 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 340 340 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 464 464 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT SEQUENCE 619 AA; 72150 MW; 839AB5D2EE545A02 CRC64;  
 SQ  
 Query Match 48.8%; Score 41; DB 1; Length 619;  
 Best Local Similarity 46.7%; Pred. No. 22;  
 Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 ATRLMKKFAEIRD 16  
 DB 88 AVLLKQLEAELSDF 102  
 RESULT 8  
 ID RPL\_TRYBB STANDARD; PRT; 1530 AA.  
 AC P08958;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE DNA-DIRECTED RNA POLYMERASE III LARGEST SUBUNIT (EC 2.7.7.6).





```
DR EMBL; AB024038; BAB02452.1; -.
DR InterPro: IPR001128; Cyt_P450.
DR Pfam: PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Transmembrane; Heme; Multigene family.
FT TRANSMEM 1 21 POTENTIAL.
FT BINDING 440 440 HEME (BY SIMILARITY).
SQ SEQUENCE 500 AA; 57080 MW; 7BD8352058F4C52C CRC64;

Query Match 47.6%; Score 40; DB 1; Length 500;
Best Local Similarity 66.7%; Pred. No. 27;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 RLMKKFAEIRD 15
      |:|:|:|:|:|:|
Db 322 RVMKKVQAEIRN 333

RESULT 11
C7C3_MAIZE STANDARD; PRT; 535 AA.
AC P93703; Q43256;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE CYTOCHROME P450 71C3 (EC 1.14.-.-).
DE CYP71C3.
GN Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE OF 2-535 FROM N.A.
RC STRAIN=CV. CI31A;
RA Frey M., Kliehm R., Siedler H., Gierl A.;
RT "Expression of a cytochrome P450 gene family in maize.";
RM Mol. Gen. Genet. 246:100-109(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CI31A;
RA Gierl A.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X81830; CAA57424.2; ALT_SEQ.
DR EMBL; Y11403; CAA72207.1; -.
DR InterPro: IPR001128; Cyt_P450.
DR Pfam: PF00067; P450; 1.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Transmembrane; Heme.
FT TRANSMEM 23 43 POTENTIAL.
FT BINDING 475 475 HEME (BY SIMILARITY).
SQ SEQUENCE 535 AA; 60715 MW; F92A696108E2ADAF CRC64;

Query Match 47.6%; Score 40; DB 1; Length 535;
Best Local Similarity 53.8%; Pred. No. 28;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 4 RLMKKFAEIRD 16
      |:|:|:|:|:|:|
```

```
Db 351 RVMKKLQAEVRAY 363

RESULT 12
VP4_ROTHT STANDARD; PRT; 774 AA.
AC P11200;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE OUTER CAPSID PROTEIN VP4 (HEMAGGLUTININ) (OUTER LAYER PROTEIN VP4)
DE [CONTAINS: OUTER CAPSID PROTEINS VP5 AND VP8].
GN S4.
OC Human rotavirus (serotype 4 / strain St. Thomas 3).
OX Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX NCBI_TaxID=10960;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=88275070; PubMed=2839714;
RA Gorziglia M., Green K., Nishikawa K., Taniguchi K., Jones R.,
RA Kapikian A.Z., Chanock R.M.;
RT "Sequence of the fourth gene of human rotaviruses recovered from
RT asymptomatic or symptomatic infections.";
RL J. Virol. 62:2978-2984(1988).
RN [2]
RP SEQUENCE OF 1-280.
RX MEDLINE=86313706; PubMed=3018754;
RA Flores J., Kapikian A.Z., Chanock R.M.;
RT "Conservation of amino acid sequence of VP8 and cleavage region of
RT 84-kDa outer capsid protein among rotaviruses recovered from
RT asymptomatic neonatal infection.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:7039-7043(1986).
CC -1- SUBCELLULAR LOCATION: OUTER CAPSID.
CC -1- PTM: VP8 IS ONE OF TWO TRYPSIN CLEAVAGE PRODUCTS OF VP4; THE
CC OTHER PRODUCT IS VP5.
CC -1- SIMILARITY: BELONGS TO THE ROTAVIRUS VP4 PROTEIN FAMILY.
DR PIR; H28839; VPRWRT.
DR InterPro: IPR000416; Cap_VP4.
DR Pfam: PF00426; VP4; 1.
KW Coat protein; Glycoprotein.
FT CHAIN 1 774 OUTER CAPSID PROTEIN VP4.
FT CHAIN 247 774 OUTER CAPSID PROTEIN VP5.
FT CARBOHYD 32 32 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 111 111 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 132 132 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 277 277 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 583 583 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 606 606 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 774 AA; 87455 MW; D397E5C6125A8FA8 CRC64;

Query Match 47.6%; Score 40; DB 1; Length 774;
Best Local Similarity 70.0%; Pred. No. 42;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 LATRLMKKFK 10
      |:|:|:|:|:|
Db 543 MATKVMKKFK 552

RESULT 13
VP4_ROTHT STANDARD; PRT; 775 AA.
AC P11197;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
```



```

RT mitochondria."
RL Nature 396:133-140(1998).
CC -!- FUNCTION: TRANSFERS THE N-ACYL DIGLYCERIDE GROUP ON WHAT WILL
CC BECOME THE N-TERMINAL CYSTEINE OF MEMBRANE LIPOPROTEINS
CC (BY SIMILARITY).
CC -!- PATHWAY: FIRST STEP IN LIPOPROTEIN BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE LGT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL-outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ235270; CAA14517.1; -.
DR InterPro; IPR001640; LGT.
DR Pfam; PF01790; LGT; 1.
DR PROSITE; PS01311; LGT; 1.
KW Transferase; Transmembrane; Complete proteome.
FT TRANSMEM 18 38 POTENTIAL.
FT TRANSMEM 64 84 POTENTIAL.
FT TRANSMEM 101 121 POTENTIAL.
FT TRANSMEM 126 146 POTENTIAL.
FT TRANSMEM 181 201 POTENTIAL.
FT TRANSMEM 210 230 POTENTIAL.
FT TRANSMEM 237 257 POTENTIAL.
SQ SEQUENCE 268 AA; 30556 MW; 451588ABF12F58A3 CRC64;

```

Query Match 46.4%; Score 39; DB 1; Length 268;  
 Best Local Similarity 58.3%; Pred. No. 20;  
 Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ATRLMKKFKAEI 13.  
 ||::||| ||  
 Db 45 ATKIERFKPEI 56

Search completed: January 29, 2002, 11:16:19  
 Job time: 148 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: January 29, 2002, 11:15:52 ; Search time 62.77 Seconds  
(without alignments)  
39.615 Million cell updates/sec

Title: US-09-763-397A-23

Perfect score: 84  
Sequence: 1 LATRLMKKFAEIRDF 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_17:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organalle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rudent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	84	100.0	153	5 Q25763	Q25763 plasmodium
2	84	100.0	161	5 Q25764	Q25764 plasmodium
3	84	100.0	163	5 Q27330	Q27330 plasmodium
4	84	100.0	782	5 Q25730	Q25730 plasmodium
5	84	100.0	782	5 Q25875	Q25875 plasmodium
6	84	100.0	782	5 Q26007	Q26007 plasmodium
7	84	100.0	782	5 Q9U431	Q9U431 plasmodium
8	84	100.0	782	5 Q9U430	Q9U430 plasmodium
9	84	100.0	782	5 Q9U429	Q9U429 plasmodium
10	84	100.0	782	5 Q9U414	Q9U414 plasmodium
11	81	96.4	782	5 Q26104	Q26104 plasmodium
12	44	52.4	1089	10 Q9LUS2	Q9LUS2 arabidopsis
13	43	51.2	353	10 Q9SB32	Q9SB32 arabidopsis
14	43	51.2	649	5 Q9VEU6	Q9VEU6 drosophila
15	42	50.0	398	10 Q9XEC5	Q9XEC5 arabidopsis
16	42	50.0	456	1 Q58576	Q58576 pyrococcus
17	42	50.0	534	12 Q65856	Q65856 beet yellow
18	42	50.0	743	10 Q9ZWA0	Q9ZWA0 arabidopsis
19	42	50.0	797	5 Q9NKN3	Q9NKN3 leishmania

#### ALIGNMENTS

RESULT 1

Q25763 PRELIMINARY; PRT; 153 AA.  
AC Q25763;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel.. 01, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE STRAIN INDJ-1 RHOPTRY-ASSOCIATED PROTEIN 1 (RAP-1) (FRAGMENT).  
GN RAP-1.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-INDJ-1 (CLONE 1 OF INDIAN ISOLATE J);  
RX MEDLINE=92244303; PubMed=1574089;  
RA Howard R.F.; Peterson C.;  
RT "The sequence of the p82 rhoptry protein is highly conserved between two Plasmodium falciparum isolates."  
RL Mol. Biochem. Parasitol. 51:327-330(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-INDJ-1 (CLONE 1 OF INDIAN ISOLATE J);  
RX MEDLINE=96379224; PubMed=8784775;  
RA Howard R.F.; Peterson C.;  
RT "Limited RAP-1 sequence diversity in field isolates of Plasmodium falciparum";  
RL Mol. Biochem. Parasitol. 77:95-98(1996).  
DR EMBL; U41075; AAC47091.1; -.  
FT NON\_TER 1 153  
SQ SEQUENCE 153 AA; 18273 MW; 6AA7AC3C4AAC447C CRC64;

Query Match 100.0%; Score 84; DB 5; Length 153;  
Best Local Similarity 100.0%; Pred. No. 4.1e-06;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LATRLMKKFAEIRDF 17

|||||

119 LATRLMKKFAEIRDF 135

|||||

```
RESULT 2
Q25764 PRELIMINARY; PRT; 161 AA.
AC Q25764;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE STRAIN INDJ-2 RHOPTRY-ASSOCIATED PROTEIN 1 (RAP-1) (FRAGMENT).
GN RAP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INDJ-2 (CLONE 2 OF INDIAN ISOLATE J);
RX MEDLINE=96379224; PubMed=1574089;
EX Howard R.F.;
RA "The sequence of the p82 rhoptry protein is highly conserved between
RT two Plasmodium falciparum isolates.";
RL Mol. Biochem. Parasitol. 51:327-330(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=INDJ-2 (CLONE 2 OF INDIAN ISOLATE J);
RX MEDLINE=96379224; PubMed=8784775;
RA Howard R.F., Peterson C.;
RT "Limited RAP-1 sequence diversity in field isolates of Plasmodium
RT falciparum.";
RL Mol. Biochem. Parasitol. 77:95-98(1996).
DR EMBL; U41076; AAC47092.1; -.
FT NON_TER 1
FT NON_TER 161
SQ SEQUENCE 161 AA; 19271 MW; 2433BF4A2FEE2E72 CRC64;

Query Match 100.0%; Score 84; DB 5; Length 161;
Best Local Similarity 100.0%; Pred. No. 4.4e-06;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LATRLMKKFKAEIRDFF 17
| | | | | | | | | | | | | | | | |
DB 122 LATRLMKKFKAEIRDFF 138

RESULT 3
Q27330 PRELIMINARY; PRT; 163 AA.
AC Q27330;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE STRAIN INDJ-2 RHOPTRY-ASSOCIATED PROTEIN 1 (RAP-1) (FRAGMENT).
GN RAP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INDC (INDIAN ISOLATE C);
RX MEDLINE=92244303; PubMed=1574089;
RA Howard R.F.;
RA "The sequence of the p82 rhoptry protein is highly conserved between
RT two Plasmodium falciparum isolates.";
RL Mol. Biochem. Parasitol. 51:327-330(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=INDC (INDIAN ISOLATE C);
RX MEDLINE=96379224; PubMed=8784775;
RA Howard R.F., Peterson C.;
RT "Limited RAP-1 sequence diversity in field isolates of Plasmodium
RT falciparum.";
RL Mol. Biochem. Parasitol. 77:95-98(1996).
DR EMBL; U41077; AAC47093.1; -.
RT
```

```
DR EMBL; U41073; AAC47089.1; -.
FT NON_TER 1
FT NON_TER 163
SQ SEQUENCE 163 AA; 19522 MW; BD43E37992DF3730 CRC64;

Query Match 100.0%; Score 84; DB 5; Length 163;
Best Local Similarity 100.0%; Pred. No. 4.4e-06;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LATRLMKKFKAEIRDFF 17
| | | | | | | | | | | | | | | | |
DB 124 LATRLMKKFKAEIRDFF 140

RESULT 4
Q25730 PRELIMINARY; PRT; 782 AA.
ID Q25730 PRELIMINARY; PRT; 782 AA.
AC Q25730;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE RHOPTRY ASSOCIATED PROTEIN-1.
GN RAP-1.
OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FC27;
RA Shi Y., Lai A.A.;
RA Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
RL EMBL; U20985; AAA63681.1; -.
SQ SEQUENCE 782 AA; 90023 MW; F69E26A2A564C8EA CRC64;

Query Match 100.0%; Score 84; DB 5; Length 782;
Best Local Similarity 100.0%; Pred. No. 2.2e-05;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LATRLMKKFKAEIRDFF 17
| | | | | | | | | | | | | | | | |
DB 736 LATRLMKKFKAEIRDFF 752

RESULT 5
Q25875 PRELIMINARY; PRT; 782 AA.
ID Q25875 PRELIMINARY; PRT; 782 AA.
AC Q25875;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE PR86 RHOPTRY PRECURSOR PROTEIN.
GN P82.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HONDURAS I CDC;
RX MEDLINE=90348711; PubMed=2200961;
RA Ridley R.G., Takacs B., Lahm H.W., Delves C.J., Goman M., Certa V.,
RA Matile H., Woollett G.R., Scaife J.G.;
RA "Characterisation and sequence of a protective rhoptry antigen from
RT Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 41:125-134(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HONDURAS I CDC;
RX MEDLINE=92244303; PubMed=1574089;
RA Howard R.F.;
RA "The sequence of the p82 rhoptry protein is highly conserved between
RT two Plasmodium falciparum isolates.";
RT
```



Mon Feb 4 15:23:45 2002

```

RN SEQUENCE FROM N.A.
RP STRAIN=COLUMBIA;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=COLUMBIA;
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT clones.";
RL DNA Res. 7:131-135(2000).
DR EMBL; AB022217; BAB02753.1; -.
KW Envelope protein.
SQ SEQUENCE 1089 AA; 119917 MW; 2ED2EAB60E409C7E CRC64;

Query Match 52.4%; Score 44; DB 10; Length 1089;
Best Local Similarity 66.7%; Pred. No. 91;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 RLMKFKAEIRD 15
DB 784 KLLKFAEIKD 795

RESULT 13
Q9SB32 PRELIMINARY; PRT; 353 AA.
AC Q9SB32;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE SRG1-LIKE PROTEIN.
GN F24A6.150 OR AT4G25310.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Wedler H., Wedler E., Wambutt R., Hoheisel J., Mewes H.W.,
RA Mayer K.F.X., Schaefer C.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP EU Arabidopsis sequencing project;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RP Wedler H., Wedler E., Wambutt R., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[4]
RN SEQUENCE OF 216-353 FROM N.A.
RA Rose M., Hempel S., Entian K.-D., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[5]
RN SEQUENCE FROM N.A.
RP EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL035396; CAA23072.1; -.
DR EMBL; AL161563; CAB81342.1; -.
DR InterPro; IPR002419; Fe_asc_oxidore.
DR InterPro; IPR002283; IPN_Synth.
DR Pfam; PF00671; Fe_Asc_Oxidore; 1.
DR PRINTS; PR00682; IPNSYNTHASE.
SQ SEQUENCE 353 AA; 39933 MW; 5D61270C34721034 CRC64;

Query Match 51.2%; Score 43; DB 10; Length 353;

QY 1 LATRLMKKFAEIRDF 17
DB 736 LATRLMKKFAEIRDF 752

RESULT 11
Q26104 PRELIMINARY; PRT; 782 AA.
AC Q26104;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE RHOPTRY ASSOCIATED PROTEIN 1.
GN RAP-1.
OS Plasmodium reichenowi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5854;
RN [1]
RP SEQUENCE FROM N.A.
RA Shi Y., Lai A.A.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20986; AAA63682.1; -.
SQ SEQUENCE 782 AA; 90160 MW; DE6D1BE2FAC308A9 CRC64;

Query Match 96.4%; Score 81; DB 5; Length 782;
Best Local Similarity 94.1%; Pred. No. 6.6e-05;
Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LATRLMKKFAEIRDF 17
DB 736 LATRLMKKFAEIRDF 752

RESULT 12
Q9LU52 PRELIMINARY; PRT; 1089 AA.
AC Q9LU52;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CHLOROPLAST OUTER ENVELOPE PROTEIN-LIKE.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
```

```

RN SEQUENCE FROM N.A.
RP STRAIN=COLUMBIA;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=COLUMBIA;
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT clones.";
RL DNA Res. 7:131-135(2000).
DR EMBL; AB022217; BAB02753.1; -.
KW Envelope protein.
SQ SEQUENCE 1089 AA; 119917 MW; 2ED2EAB60E409C7E CRC64;

Query Match 52.4%; Score 44; DB 10; Length 1089;
Best Local Similarity 66.7%; Pred. No. 91;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 RLMKFKAEIRD 15
DB 784 KLLKFAEIKD 795

RESULT 13
Q9SB32 PRELIMINARY; PRT; 353 AA.
AC Q9SB32;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE SRG1-LIKE PROTEIN.
GN F24A6.150 OR AT4G25310.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Wedler H., Wedler E., Wambutt R., Hoheisel J., Mewes H.W.,
RA Mayer K.F.X., Schaefer C.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP EU Arabidopsis sequencing project;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RP Wedler H., Wedler E., Wambutt R., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[4]
RN SEQUENCE OF 216-353 FROM N.A.
RA Rose M., Hempel S., Entian K.-D., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[5]
RN SEQUENCE FROM N.A.
RP EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL035396; CAA23072.1; -.
DR EMBL; AL161563; CAB81342.1; -.
DR InterPro; IPR002419; Fe_asc_oxidore.
DR InterPro; IPR002283; IPN_Synth.
DR Pfam; PF00671; Fe_Asc_Oxidore; 1.
DR PRINTS; PR00682; IPNSYNTHASE.
SQ SEQUENCE 353 AA; 39933 MW; 5D61270C34721034 CRC64;

Query Match 51.2%; Score 43; DB 10; Length 353;

QY 4 RLMKFKAEIRD 15
DB 784 KLLKFAEIKD 795
```



```
Best Local Similarity 58.3%; Pred. No. 42;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 6 MKKFAEIRDF 17
Db 92 LDKESDIOFF 103

RESULT 14
Q9VEU6 PRELIMINARY; PRT; 649 AA.
AC Q9VEU6
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CG14885 PROTEIN.
GN CG14885.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S.G., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Morkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Rao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AF003713; AAF55322.1; -.
DR HSP: P16068; IAWN.
DR FlyBase: FBgn0038435; CG14885.
DR InterPro: IPR001054; Guanylt_cyclase.
DR Pfam: PF00211; guanylate_cyc; 1.
DR SMART: SM00044; CYCC; 1.
DR PROSITE: PS00452; GUANYLATE_CYCLASES_1; 1.
DR PROSITE: PS0125; GUANYLATE_CYCLASES_2; 1.
KW Lyase.
```

```
SQ SEQUENCE 649 AA; 73892 MW; 3CF9F52B7E5D7505 CRC64;

Query Match 51.2%; Score 43; DB 5; Length 649;
Best Local Similarity 81.8%; Pred. No. 78;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LATRLMKKFKKA 11
Db 546 LALRVNKKFKA 556

RESULT 15
Q9XEC5 PRELIMINARY; PRT; 398 AA.
AC Q9XEC5;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE HYPOTHETICAL 44.0 KDA PROTEIN.
GN T26N6.8 OR AT4G04480.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Zhong J., Ma P., Parnell L.D., Chen C.-N., Chen E.Y.;
RT "Genomic sequence of Arabidopsis thaliana BAC T26N6, chromosome IV,
RT 19.3 cm."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Zhong J., Ma P., Parnell L.D., Chen C.N., Chen E.Y.; Mewes H.W.,
RA Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF076243; AAD29760.1; -.
DR EMBL: AL161500; CAB77916.1; -.
KW Hypothetical protein.
SQ SEQUENCE 398 AA; 43961 MW; 0077BE2298E94AD7 CRC64;

Query Match 50.0%; Score 42; DB 10; Length 398;
Best Local Similarity 60.0%; Pred. No. 70;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 ATRLMKKFAEIRDF 16
Db 165 ATRLFKKLRVEIEMF 179

Search completed: January 29, 2002, 11:15:52
Job time: 212 sec
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 10:59:19 ; Search time 66.28 Seconds  
(without alignments)  
16.764 Million cell updates/sec

Title: US-09-763-397A-24  
Perfect score: 74  
Sequence: 1 QYIKANSKFIGITEL 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues  
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_1101.\*  
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	74	100.0	15	11	AA06310		Tetanus toxin epit
2	74	100.0	15	18	AAW35506		Universal T-cell e
3	74	100.0	15	18	AAW11505		Tetanus toxoid uni
4	74	100.0	15	19	AAW67033		Tetanus toxin frag
5	74	100.0	15	19	AAW71321		Universal helper T
6	74	100.0	15	20	AAW04051		T-Helper epitope f
7	74	100.0	15	20	AAW67578		T-cell epitope pep
8	74	100.0	15	20	AAW73220		Tetanus toxoid epi
9	74	100.0	15	21	AAW45511		Tetanus P2 epitope
10	74	100.0	15	21	AAW82637		Tetanus toxoid T c
11	74	100.0	15	21	AAW92625		Foreign epitope P2

12	74	100.0	15	21	AAW84427	Amino acid sequenc
13	74	100.0	15	21	AAW70300	Clostridium tetani
14	74	100.0	15	21	AAW44763	Tetanus toxoid pro
15	74	100.0	15	22	AAW85451	Wild-type TT830 (t
16	74	100.0	15	22	AAW61956	Tetanus toxoid uni
17	74	100.0	15	22	AAW20143	Tetanus toxin T-ce
18	74	100.0	15	22	AAW68636	HER-2 B cell pepti
19	74	100.0	15	22	AAW46172	Tetanus toxoid TT8
20	74	100.0	15	22	AAW49071	Tetanus toxoid TT
21	74	100.0	16	18	AAW35445	T-cell stimulatory
22	74	100.0	16	20	AAW23705	Clostridium tetani
23	74	100.0	17	15	AAW62692	Helper T cell epit
24	74	100.0	17	16	AAW82573	Tetanus toxin help
25	74	100.0	17	17	AAW05599	Tetanus toxin help
26	74	100.0	17	17	AAW88395	T-cell antigen TT2
27	74	100.0	17	21	AAW99274	HLA class II bindi
28	74	100.0	17	21	AAW80056	Pathogen derived T
29	74	100.0	17	21	AAW54539	T helper cell (Th)
30	74	100.0	17	21	AAW58768	Unidentified pepti
31	74	100.0	17	22	AAW62904	Amino acid residue
32	74	100.0	17	22	AAW84435	Amino acid sequenc
33	74	100.0	17	22	AAW30941	Antigenic fragment
34	74	100.0	17	22	AAW31029	Antigenic fragment
35	74	100.0	17	22	AAW31118	Antigenic fragment
36	74	100.0	17	22	AAW15589	Peptide 5 for pept
37	74	100.0	18	20	AAW26607	HIV-derived lipope
38	74	100.0	19	21	AAW99055	HLA class II bindi
39	74	100.0	22	22	AAW46175	Tetanus toxoid 830
40	74	100.0	22	22	AAW46178	Tetanus toxoid 830
41	74	100.0	22	22	AAW46196	Tetanus toxoid epi
42	74	100.0	22	22	AAW46203	Human APP A-beta 1
43	74	100.0	25	21	AAW92650	PSMpep007 - P2 ins
44	74	100.0	25	21	AAW92651	PSMpep008 - P2 ins
45	74	100.0	25	21	AAW92652	PSMpep009 - P2 ins

## ALIGNMENTS

RESULT 1	
AA06310	
ID	AA06310 standard; protein; 15 AA.
XX	
AC	AA06310;
XX	
DT	04-DEC-1990 (first entry)
XX	
DE	Tetanus toxin epitope.
XX	
KW	Tetanus toxin; vaccine; major histocompatibility complex; MHC;
KW	antimalarial.
XX	
OS	Synthetic.
XX	
PN	EP378881-A.
XX	
PD	25-JUL-1990.
XX	
PF	27-DEC-1989; 89EP-0203318.
XX	
PR	16-NOV-1989; 89IT-0022409.
PR	17-JAN-1989; 89IT-0019110.
XX	(ENTE ) ENRICALCHE SPA.
PA	
XX	
PI	Pessi A, Bianchi E, Verdini AS, Corradin G;
XX	
DR	WPI; 1990-225582/30.
XX	
PT	Synthetic peptide(s) corresp. to tetanus toxin epitope(s) - used
PT	as universal carriers for prepn. of immunogenic conjugate(s) for
XX	use as vaccines.

PS Claim 1; Page 17; 20pp; English.

XX Epitopic peptides may be used with synthetic hapten derived from  
CC a pathogen to generate an immune response to the pathogen.  
CC Peptides are recognised by numerous T-helper cell clones within  
CC the context of a wide range of alleles of the human MHC.  
CC The peptides may be used in an antimalarial vaccine inducing Ab.  
CC response to P.falciparum.  
XX  
SQ Sequence 15 AA;

Query Match 100.0%; Score 74; DB 11; Length 15;  
Best Local Similarity 100.0%; Pred. No. 2.3e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QYIKANSKFIGITEL 15  
Db 1 qyikanskfigitel 15  
|||||

RESULT 2  
AAW35506  
ID AAW35506 standard; peptide; 15 AA.  
XX  
AC AAW35506;

DT 22-APR-1998 (first entry)

DE Universal T-cell epitope peptide SEQ ID NO:8.

KW T-cell stimulatory peptide; immunogen; non-dendritic; carrier; tumour;  
KW scaffold; inhibition; metastasis; wound healing; solid phase.  
XX  
OS Unidentified.

PN W09738011-A1.

PD 16-OCT-1997.

PF 03-APR-1997; 97WO-DE00146.

PR 03-APR-1996; 96DK-0000398.

PA (PEPR-) PEPRESEARCH AS.

PI Heegaard PMH, Jakobsen PH;

DR WPI; 1997-512645/47.

XX Non-dendritic peptide carrier linked to a solid phase - useful as a  
PT diagnostic agent and as a scaffold for production of chemical  
PT derivatives

PS Example 20; Page 124; 262pp; English.

XX A non-dendritic peptide carrier (A) has been developed which is coupled  
CC through a linker to a solid phase, forming a complex of (A)-solid phase.  
CC Where (A) comprises 10-50 amino acids capable of forming a secondary  
CC structure in a benign buffer after liberation from the solid phase, and  
CC further the (A)-solid phase complex comprises an immunogenic substance  
CC and/or an immune mediator coupled on (A). The present sequence  
CC represents a peptide used in an example from the present invention. An  
CC (A)-solid phase complex can be used as a scaffold for the production of  
CC chemical derivatives, characterised by covalently attaching molecules at  
CC attachment points. Alternatively (A) is used as a scaffold-peptide for  
CC the incorporation into an Immunostimulating Complex (Iscom) resulting an  
CC (A)-Iscom complex which is used for the chemical coupling of antigenic  
CC substances in an aqueous solution by conjugation. (A) derivatised with  
CC one or more peptides having fibronectin-, laminin- or vitronectin-like  
CC binding activities can be used for the promotion of cell-attachment to  
CC plastic surfaces, in particular to inhibit tumour growth and metastasis,  
CC and for promotion of wound healing. Also a derivatised (A) can be used

CC for the selection of specifically-binding aptamers or as a diagnostic  
CC agent. Such diagnostic-(A) molecules could be used to detect molecules  
CC derived from or indicative of pregnancy or of a disease, such as an  
CC infectious, autoimmune or cancerous disease.  
XX  
SQ Sequence 15 AA;

Query Match 100.0%; Score 74; DB 18; Length 15;  
Best Local Similarity 100.0%; Pred. No. 2.3e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QYIKANSKFIGITEL 15  
Db 1 qyikanskfigitel 15  
|||||

RESULT 3  
AAW11505  
ID AAW11505 standard; Protein; 15 AA.  
XX  
AC AAW11505;

DT 24-SEP-1997 (first entry)

DE Tetanus toxoid universal Th epitope TT830.

KW Humanised antibody; anti-Fc receptor; H22; bifunctional; bispecific;  
KW fusion protein; chimera; tetanus toxoid; helper T cell epitope;  
KW antigen presentation; ds.  
XX  
OS Clostridium tetani.

PN W09640789-A1.

PD 19-DEC-1996.

PF 07-JUN-1996; 96WO-US09988.

PR 07-JUN-1995; 95US-0484172.

PA (MEDA-) MEDAREX INC.

PI Deo YM, Goldstein J, Graziano R, Somasundaram C;

DR WPI; 1997-052242/05.

DR N-PSDB; AAT58127.

XX Recombinant, multi-specific anti-Fc receptor antibody molecules -  
PT also comprise an anti-target portion, used for the treatment of  
PT cancer, autoimmune disease and pathogenic infection

PS Example 7; Fig 24; 115pp; English.

XX Synthetic DNA coding for the wild-type universal Th epitope from  
CC tetanus toxoid, designated TT830, was fused to the 3'-end of DNA  
CC encoding heavy chain sequences from the humanised anti-Fc gamma RI  
CC monoclonal antibody H22. The resulting fusion protein was shown to  
CC be significantly more efficient in antigen presentation and T cell  
CC stimulation than the TT830 epitope alone. A similar fusion  
CC construct was prepared coding for a mutant, antagonistic form of the  
CC epitope (designated TT833S) fused to the anti-Fc gamma RI. The  
CC Fab22-TT833S is at least 100 times more effective than TT833S in  
CC inhibiting T cell activation.

SQ Sequence 15 AA;

Query Match 100.0%; Score 74; DB 18; Length 15;  
Best Local Similarity 100.0%; Pred. No. 2.3e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QYIKANSKFIGITEL 15

Db 1 qvikanskfigitel 15  
|||||

## RESULT 4

AAW67033  
ID AAW67033 standard; peptide; 15 AA.

XX AAW67033;  
XX

DT 15-DEC-1998 (first entry)  
XX

DE Tetanus toxin fragment (residues 830-844).  
XX

KW Tetanus toxin; vaccine; antibody; carbohydrate peptide conjugate;  
KW dendrimeric poly-lysine; epitope; tumour.  
XX

OS Clostridium tetani.  
XX

PN W09843677-A1.  
XX

PD 08-OCT-1998.  
XX

PF 27-MAR-1998; 98WO-EP01922.  
XX

PR 27-MAR-1997; 97US-0041726.  
XX

PA (INSP ) INST PASTEUR.  
XX

PI Bay S, Cantacuzene D, Leclerc C, Lo-man R;  
XX

DR WPI; 1998-557071/47.  
XX

PT Carbohydrate peptide conjugate used as vaccine - comprises carrier  
PT with dendrimeric poly-lysine enabling multiple epitopes to be  
PT covalently attached  
XX

PS Disclosure: Page 13; 55pp; English.  
XX

CC The invention relates to a new carbohydrate peptide conjugate, which  
CC comprises a carrier with a dendrimeric poly-lysine enabling multiple  
CC epitopes to be covalently attached to it. Also claimed are: (1) an  
CC antibody purified from biological fluid or cells of organisms  
CC administered with the carbohydrate peptide conjugate, and (2) a diagnosis  
CC kit comprising antigen-specific antibodies elicited by immunisation with  
CC the carbohydrate peptide conjugate. The peptide conjugate, antibody and  
CC diagnosis kit are used to provide pharmaceutical compositions and  
CC vaccines against tumours. These can be used to support an immune response  
CC against viral infections caused by hepatitis virus, HIV or cytomegalo  
CC virus. They can be used to enhance immune responses, especially B- and T-  
CC cell responses, of humans and animals against bacterial infections. The  
CC carbohydrate peptide conjugate stimulates the antibody and T-cell  
CC response without stimulating undesired immune responses. The composition  
CC is capable of increasing the survival of tumour bearing humans and  
CC animals. The present sequence corresponds to residues 830-844 of tetanus  
CC toxin. The synthetic peptide corresponding to this sequence may be used  
CC as an epitope in a carbohydrate peptide conjugate.  
XX

SQ Sequence 15 AA;

Query Match 100.0%; Score 74; DB 19; Length 15;  
Best Local Similarity 100.0%; Pred. No. 2.3e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QVIKANSKFIGITEL 15

Db 1 qvikanskfigitel 15  
|||||

## RESULT 5

AAW71321  
ID AAW71321 standard; peptide; 15 AA.

XX AAW71321;  
AC

DT 26-NOV-1998 (first entry)  
XX

DE Universal helper T-cell epitope P2 derived from tetanus toxin.  
XX

KW Liver stage; Plasmodium; Navy Yoelii Liver Stage 3 antibody; NYLS3;  
KW hepatic and erythrocytic stage protein; PyHep17; vaccine;  
KW malaria parasite; teanus toxin; P2; helper T-cell epitope.  
XX

OS Synthetic.  
OS Clostridium tetani.

PN US5814617-A.  
XX

PD 29-SEP-1998.  
XX

PF 07-OCT-1994; 94US-0319704.  
XX

PR 07-OCT-1994; 94US-0319704.  
XX

PA (USNA ) US SEC OF NAVY.  
XX

PI Charoenvit Y, Doolan DL, Hedstrom RC, Hoffman SL;  
XX

DR WPI; 1998-541794/46.  
XX

PT Vaccine for protecting mammal against infection by malaria caused by  
PT Plasmodium species - comprises a first nucleic acid encoding a first  
PT polypeptide capable of eliciting an immune reaction against an  
PT antigen expressed during the liver  
XX

PS Disclosure; Column 12; 24pp; English.  
XX

CC AAW71321-22 represent universal helper T-cell epitopes derived from  
CC tetanus toxin. They are used to enhance host immune response to  
CC vaccines. The specification describes a Plasmodium yoelii liver stage  
CC 17 kDa hepatic and erythrocytic stage protein designated PyHep17. This  
CC protein elicits a response from an Igi monoclonal antibody designated  
CC Navy Yoelii Liver Stage 3 (NYLS3). This antibody does not recognise  
CC sporozoites, but does recognise P. yoelii liver stage parasites. NYLS3  
CC eliminates upto 90% of liver stage parasites. The specification  
CC describes a vaccine for reducing the severity or incidence of infection  
CC by a malaria parasite of the genus Plasmodium. The DNA vaccine comprises  
CC exon 1 and part of exon 2 of the PyHep17 gene.  
XX

SQ Sequence 15 AA;

Query Match 100.0%; Score 74; DB 19; Length 15;  
Best Local Similarity 100.0%; Pred. No. 2.3e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QVIKANSKFIGITEL 15

Db 1 qvikanskfigitel 15  
|||||

## RESULT 6

AAW04051  
ID AAW04051 standard; peptide; 15 AA.

XX AAW04051;  
AC

DT 04-JAN-2000 (first entry)  
XX

DE T-Helper epitope from tetanus toxoid.  
XX

KW Covalently reactive antigen analog; CRAA; catalytic antibody;  
KW electrophilic reaction centre; phosphonate; boronate; vaccine;  
KW transition state analog; TSA; isostere; gpl20; HIV-1; T-helper;  
KW tetanus; toxoid; B-T-epitope.

XX OS Clostridium tetani.  
 XX PN WO9948925-A1.  
 XX XX  
 XX PD 30-SEP-1999.  
 XX XX  
 XX PF 23-MAR-1999; 99WO-US06325.  
 XX PR 23-MAR-1998; 98US-0046373.  
 XX PA (UYNE-) UNIV NEBRASKA.  
 XX PI Paul S, Gololobov G, Smith L;  
 XX DR WPI; 1999-591076/50.  
 XX PT New covalently reactive antigen analogs used for treating e.g.  
 XX PT autoimmune diseases, lymphoproliferative disorders, cancers, microbial  
 XX PT infections, ischemic and reperfusion injury or septic shock -  
 XX PS Disclosure; Page 86; 158pp; English.  
 XX CC The patent discloses new covalently reactive antigen analogs (CRAA)  
 XX CC of formula XI-Y-E-X2, in which XI and X2 represent peptide sequences  
 XX CC of an epitope of a disease-associated protein, Y is a positively  
 XX CC charged amino acid residue, preferably Lys or Arg, and E is an  
 XX CC electrophilic reaction centre, preferably a phosphonate or boronate  
 XX CC moiety. Depending on the identity of the epitope, the CRAA may be used  
 XX CC to stimulate production of catalytic antibodies specific for  
 XX CC predetermined antigens associated with particular medical disorders.  
 XX CC They may also be used to permanently inactivate endogenously produced  
 XX CC catalytic antibodies produced in certain autoimmune diseases as well as  
 XX CC in certain lymphoproliferative disorders.  
 XX CC Amongst the specifically exemplified CRAAs is one based on residues  
 XX CC 421-436 of a B-cell epitope of gp120 (see AAY04046) which may be used  
 XX CC to counter HIV-1 infections. When used as an immunogen, preferably this  
 XX CC CRAA is conjugated at its N-terminal to a T-helper epitope from tetanus  
 XX CC toxoid. The present sequence represents the T-helper epitope and  
 XX CC corresponds to residues 830-844 of the toxoid.  
 XX SQ Sequence 15 AA;  
 Query Match 100.0%; Score 74; DB 20; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QYIKANSKFIGITEL 15  
 DB 1 qyikanskfigitel 15  
 RESULT 7  
 AAW67578  
 ID AAW67578 standard; peptide; 15 AA.  
 XX AC  
 XX AC AAW67578;  
 XX DT 02-MAR-1999 (first entry)  
 XX DE T-cell epitope peptide #4 for chimeric fimbrin/T-cell epitope peptide.  
 XX KW Chimeric; non-typable Haemophilus influenzae; fimbrin; T-cell epitope;  
 XX KW immunogenic composition; immune response.  
 XX OS Synthetic.  
 XX PN US5843464-A.  
 XX PD 01-DEC-1998.  
 XX PF 02-JUN-1995; 95US-0460502.

XX PR 02-JUN-1995; 95US-0460502.  
 XX PA (OHIS ) UNIV OHIO STATE.  
 XX PI Bakaletz LO, Kaumaya PTP;  
 XX DR WPI; 1999-044514/04.  
 XX XX Synthetic chimeric fimbrin peptide - useful for vaccination against  
 XX PT non-typable Haemophilus influenzae  
 XX PS Disclosure; Column 4; 16pp; English.  
 XX CC The invention relates to the manufacture of a synthetic chimeric peptide  
 XX CC comprising a non-typable Haemophilus influenzae fimbrin peptide fused via  
 XX CC a linker peptide to a T-cell epitope peptide. The chimeric peptide is  
 XX CC used in immunogenic compositions which induce an immune response against  
 XX CC non-typable Haemophilus influenzae. This sequence represents an example  
 XX CC of a T-cell epitope peptide used to generate the chimeric peptide.  
 XX SQ Sequence 15 AA;  
 Query Match 100.0%; Score 74; DB 20; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QYIKANSKFIGITEL 15  
 DB 1 qyikanskfigitel 15  
 RESULT 8  
 AAW73220  
 ID AAW73220 standard; Protein; 15 AA.  
 XX AC  
 XX AC AAW73220;  
 XX DT 25-JAN-1999 (first entry)  
 XX DE Tetanus toxoid epitope.  
 XX KW Multispecific single chain antibody; antibody H22; tumour cell; therapy;  
 XX KW antibody-dependent cellular cytotoxicity; ADCC; HER 2/neu; infection;  
 XX KW epidermal growth factor receptor; breast cancer; ovarian cancer.  
 XX OS Synthetic.  
 XX PN US5837243-A.  
 XX PD 17-NOV-1998.  
 XX PF 07-JUN-1996; 96US-0661052.  
 XX PR 07-JUN-1996; 96US-0661052.  
 XX PR 07-JUN-1995; 95US-0484172.  
 XX PA (MEDA-) MEDAREX INC.  
 XX PI Deo YM, Goldstein J, Graziano R, Somasundaram C;  
 XX DR WPI; 1999-023374/02.  
 XX XX Specific killing of tumour cells - using a multi-specific molecule  
 XX PT comprising an anti-Fc receptor antibody and a portion which binds to  
 XX PT a target cell  
 XX PS Example 7; Column 27; 57pp; English.  
 XX CC This sequence represents a tetanus toxoid epitope and is recognised  
 XX CC by the multispecific single chain antibody designated H22. The  
 XX CC antibody can be used in the method of the invention for inducing

CC antibody-dependent cellular cytotoxicity (ADCC) against a tumour cell  
 CC which is characterised by overexpression of HER 2/neu or epidermal growth  
 CC factor receptor (EGFR), comprises contacting the tumour cell with a  
 CC multispecific protein molecule (preferably a single chain antibody)  
 CC comprising: (a) an anti-FC receptor antibody or an antigen binding  
 CC fragment; (b) a portion which binds to HER 2/neu; and (c) a portion which  
 CC binds to EGFR. The method can be used for treating cancers especially  
 CC breast cancer or ovarian cancer. The multispecific antibody can also  
 CC be administered prophylactically to vaccinate a subject against infection  
 CC by a target cell.

XX Sequence 15 AA;

Query Match 100.0%; Score 74; DB 20; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QYIKANSKFIGITEL 15  
 Db 1 qvikanskfigitel 15

RESULT 9  
 AAB45511  
 ID AAB45511 standard; Protein; 15 AA.

XX AAB45511;

AC AAB45511;  
 XX 26-FEB-2001 (first entry)  
 DT Tetanus P2 epitope SEQ ID NO: 23.

DE Asthma; IL-5; interleukin-5; allergy; cytokine; helminthic infection;  
 KW cancer; eosinophilia; vaccine; allergic rhinitis.

XX Clostridium tetani.

OS WO200065058-A1.

XX 02-NOV-2000.

PD 19-APR-2000; 2000WO-DK00205.

XX 23-APR-1999; 99DK-0000552.

PR 06-MAY-1999; 99US-0132811.

XX (MEBI-) M & E BIOTECH AS.

PA Klysner S;

XX WPI; 2000-672791/65.

DR Down-regulating interleukin 5 (IL-5) activity in humans by  
 XX administering IL-5 and/or an IL-5 analogue, useful in the treatment,  
 PT prophylaxis or amelioration of asthma or other chronic allergic  
 PT conditions -

XX Example 1; Page 137; 172pp; English.

XX The present invention is concerned with methods of treating asthma,  
 CC eosinophilia, allergic rhinitis and other allergic diseases. These  
 CC involve the use of interleukin-5 (IL-5) analogues and modified IL-5  
 CC proteins and their coding sequences to down-regulate IL-5 activity and  
 CC thus reduce eosinophil numbers. The allergic diseases may be treated  
 CC using autovaccines, nucleic acid vaccines or live vaccines. In addition,  
 CC it is possible that they may be used in the treatment of cancer and  
 CC helminthic infections.

XX Sequence 15 AA;

Query Match 100.0%; Score 74; DB 21; Length 15;

Best Local Similarity 100.0%; Pred. No. 2.3e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QYIKANSKFIGITEL 15  
 Db 1 qvikanskfigitel 15

RESULT 10

AA82637  
 ID AAY82637 standard; peptide; 15 AA.

XX AAY82637;

XX 07-AUG-2000 (first entry)

DE Tetanus toxoid T cell epitope peptide SEQ ID NO:13.

XX T cell epitope; B cell epitope; allergy; allergen; antigenic;  
 KW anti-allergic; antiasthmatic; antiinflammatory; dermatological;  
 KW immunosuppressive; vaccine; rhinitis; sinusitis; bronchial asthma;  
 KW atopic dermatitis; acute urticaria; chronic urticaria;  
 KW gastro-intestinal syndrome; food allergen; oro-pharyngeal syndrome;  
 KW anaphylactic reaction; drug hypersensitivity; allergic reaction.

XX Clostridium tetani.

OS Synthetic.

XX WO200006694-A2.

PN 10-FEB-2000.

XX 20-JUL-1999; 99WO-BE00092.

XX 30-JUL-1998; 98EP-0870167.

XX (UNIO ) UCB SA.

XX Saint-Remy J, Jacquemin M;

XX WPI; 2000-422470/36.

XX New compound for prevention and treatment of allergies comprises at  
 PT least one allergen antigenic determinant recognized by a B cell and at  
 PT least one antigenic determinant which does not trigger T cell  
 PT activation -

XX Example 6; Page 30; 50pp; English.

XX The present invention describes a compound (I) for the prevention and/or  
 CC treatment of allergy. The compound comprises at least one allergen  
 CC antigenic determinant (I) recognised by a B cell or an antibody secreted  
 CC by a B cell of a non-atopic individual and at least one antigenic  
 CC determinant (II) different from the allergen that triggers T cell  
 CC activation. (I) has anti-allergic, antiasthmatic, antiinflammatory,  
 CC dermatological and immunosuppressive activities, and can be used in a  
 CC vaccine. (I) may be used in a pharmaceutical or cosmetic medicament to  
 CC treat and/or prevent allergies or a disease of allergic origin,  
 CC especially hypersensitivities. These include rhinitis, sinusitis,  
 CC bronchial asthma, atopic dermatitis, some forms of acute and chronic  
 CC urticaria, gastro-intestinal syndromes associated with the ingestion of  
 CC food allergens, oro-pharyngeal syndrome, anaphylactic reactions  
 CC associated with drug hypersensitivities and/or a mixture of these. The  
 CC use of (I) in the treatment of allergic conditions avoids the need for  
 CC drug treatment, which often causes undesirable side-effects. Also, prior  
 CC art drug therapies alleviate symptoms, but do not influence their  
 CC causes, however (I) actually combats the cause of an allergic reaction.  
 CC The present sequence represents a peptide, which is used in an  
 CC example from the present invention.

XX Sequence 15 AA;

```
Query Match      100.0%; Score 74; DB 21; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QYIKANSKFIGITEL 15
Db 1 qyikanskfigitel 15
| | | | | | | | | | | | | | | |

RESULT 11
AAY92625
ID AAY92625 standard; Protein; 15 AA.
XX
AC AAY92625;
XX
DT 10-AUG-2000 (first entry)
XX
DE Foreign epitope P2.
XX
KW Foreign epitope; P2; prostate specific membrane antigen; PSM; Her2;
KW Heregulin 2; Fibroblast growth factor 8b; FGF8b; vaccination;
KW cytotoxic T-lymphocyte immunity; self-protein; cancer; breast cancer;
KW prostate cancer; cell-associated peptide antigen.
XX
OS Clostridium tetani.
XX
PN WO200020027-A2.
XX
PD 13-APR-2000.
XX
PF 05-OCT-1999; 99WO-DK00525.
XX
PR 05-OCT-1998; 98DK-0001261.
XX
PR 20-OCT-1998; 98US-0105011.
XX
PA (MEBI-) M & E BIOTECH AS.
XX
PI Steinaa L, Mouritsen S, Nielsen KG, Haanning J, Leach D, Dalum I;
PI Gautam A, Birk P, Karlsson G;
XX
DR WPI: 2000-349917/30.
XX
DR N-PSDB; AAA09460.
XX
PT Inducing immune responses to weakly immunogenic, tumor associated
PT peptide antigens for the treatment of breast and prostate cancer
XX
PS Example 1; Page 213; 220pp; English.
XX
CC The claims detail a method for inducing immune responses against weakly
CC immunogenic cell-associated peptide antigens (PA) such as those
CC associated with cancers (i.e. self-proteins), for example, human
CC prostate specific membrane antigen (PSM), heregulin 2 (Her2) and/or
CC fibroblast growth factor 8b (FGF8b). The method comprises effecting
CC simultaneous presentation by antigen producing cells (APCs) of the
CC animals immune system of: (1) at least 1 CTL (cytotoxic T-lymphocyte)
CC group derived from the PA and/or at least 1 B-cell group derived from the
CC cell-associated PA; and (2) at least 1 first T helper cell group which is
CC foreign to the animal. Analogues of human PSM, human Her2 and
CC human/murine FGF8b comprising a substantial part of all known and
CC predicted CTL and B-cell epitopes of the respective PA and including at
CC least one foreign T helper epitope (e.g. P2 and/or P30) are also claimed.
CC The method is used to treat prostate, prostate/breast or breast cancer
CC when the PA is human PSM, FGF8b and Her2, respectively.
XX
SQ Sequence 15 AA;

Query Match      100.0%; Score 74; DB 21; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QYIKANSKFIGITEL 15
Db 1 qyikanskfigitel 15
| | | | | | | | | | | | | | | |

RESULT 12
AAY84427
ID AAY84427 standard; Peptide; 15 AA.
XX
AC AAY84427;
XX
DT 25-JUL-2000 (first entry)
XX
DE Amino acid sequence of the tetanus toxoid P2 epitope.
XX
KW Osteoprotegerin ligand; OPG; osteoprotegerin; osteoclastogenesis;
KW tumour necrosis factor receptor; type II transmembrane protein;
KW osteoclast differentiation; CSF-1; osteoclast activator;
KW immune response; osteoporosis; bone resorption;
KW tetanus toxoid P2 epitope.
XX
OS Clostridium tetani.
XX
PN WO200015807-A1.
XX
PD 23-MAR-2000.
XX
PF 13-SEP-1999; 99WO-DK00481.
XX
PR 15-SEP-1998; 98DK-0001164.
XX
PR 02-OCT-1998; 98US-0102896.
XX
PA (MEBI-) M & E BIOTECH AS.
XX
PI Halkier T, Haanning J;
XX
DR WPI: 2000-271444/23.
XX
PT In vivo down-regulation of osteoprotegerin ligand (OPGL) activity used
PT to treat, prevent and ameliorate osteoporosis -
XX
PS Example; Page 106; 110pp; English.
XX
CC The present sequence represents the tetanus toxoid P2 epitope. It is
CC used to create a fusion protein with murine osteoprotegerin ligand
CC (OPGL). Osteoprotegerin is a secreted member of the tumour necrosis
CC factor receptor family, which blocks osteoclastogenesis in a dose
CC dependent manner. The OPGL protein is synthesised as a type II
CC transmembrane protein. The murine and human OPGL polypeptides are 87%
CC homologous. OPGL is a potent osteoclast differentiation factor when
CC combined with CSF-1. It is not capable of inducing osteoclast
CC differentiation in the absence of CSF-1. OPGL is also an activator of
CC mature osteoclasts. The specification describes a method for the in vivo
CC down-regulation of OPGL activity in an animal. The method comprises
CC using at least one OPGL polypeptide or subsequence, and/or at least one
CC OPGL analogue to induce an immune response in the animal. The method
CC and OPGL polypeptide are useful for treating, preventing and ameliorating
CC osteoporosis or other diseases or conditions characterised by excessive
CC bone resorption.
XX
SQ Sequence 15 AA;

Query Match      100.0%; Score 74; DB 21; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QYIKANSKFIGITEL 15
Db 1 qyikanskfigitel 15
| | | | | | | | | | | | | | | |

RESULT 13
AAY70300
ID AAY70300 standard; peptide; 15 AA.
```



```

XX AC AAY70300;
XX
XX DT 06-JUN-2000 (first entry)
XX
XX DE Clostridium tetani tetanus toxoid T-cell epitope, P589.
XX
XX KW Recombinant protein; CDC/NIIMALVAC-1; multivalent; malaria; vaccine;
XX T-cell epitope; tetanus toxoid; antigenic epitope; treatment;
XX KW circumsporozoite protein; CSP; sporozoite surface protein-2; SSP-2;
XX KW liver stage antigen-1; LSA-1; merozoite surface protein-1; MSP-1; MSP-2;
XX KW apical membrane antigen-1; AMA-1; erythrocyte binding antigen-175;
XX KW EBA-175; rhoptry associated protein-1; RAP-1; gamete specific antigen;
XX KW Pfg27; antiparasitic; prevention; anti-CDC/NIIMALVAC-1 antibody.
XX
XX OS Clostridium tetani.
XX
XX PN WO200011179-A1.
XX
XX PD 02-MAR-2000.
XX
XX PF 19-AUG-1999; 99WO-US18869.
XX
XX PR 21-AUG-1998; 98US-0097703.
XX
XX PA (NAIM-) NAT INST IMMUNOLOGY.
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX PI Lal AA, Shi YP, Hasnain SE;
XX
XX DR WPI; 2000-237654/20.
XX
XX PT Novel recombinant protein as vaccine for treating malarial infection
XX PT comprises antigenic peptides obtained from different stages of
XX PT plasmodium falciparum life cycle -
XX
XX PS Claim 2; Page 17; 52pp; English.
XX
XX CC The present sequence is the tetanus toxoid T-cell epitope P589, derived
XX CC from Clostridium tetani. It is used in the construction of recombinant
XX CC protein CDC/NIIMALVAC-1, which is a multivalent, multistage malarial
XX CC vaccine. The recombinant protein comprises, melittin signal peptide,
XX CC (His)6 tag, T-cell epitope from tetanus toxoid and 21 antigenic epitopes
XX CC from circumsporozoite protein (CSP), sporozoite surface protein-2
XX CC (SSP-2), liver stage antigen-1 (LSA-1), merozoite surface protein-1
XX CC (MSP-1), MSP-2, apical membrane antigen-1 (AMA-1), erythrocyte binding
XX CC antigen-175 (EBA-175), rhoptry associated protein-1 (RAP-1) and gamete
XX CC specific antigen, Pfg27. These epitopes were obtained at different stages
XX CC of the life cycle of P. falciparum. CDC/NIIMALVAC-1 vaccine has
XX CC antiparasitic activity and can be used for treatment and prevention of
XX CC malarial infections. Anti-CDC/NIIMALVAC-1 antibodies can be used for
XX CC detecting P. falciparum in biological samples.
XX
XX SQ Sequence 15 AA;
XX
XX Query Match 100.0%; Score 74; DB 21; Length 15;
XX Best Local Similarity 100.0%; Pred. No. 2.3e-07;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 QYIKANSKFIGITEL 15
XX | | | | | | | | | |
XX Db 1 qyikanskfigitel 15
XX
XX RESULT 14
XX AAY44763
XX ID AAY44763 standard; peptide; 15 AA.
XX
XX AC AAY44763;
XX
XX DT 04-MAY-2000 (first entry)
XX
XX KW HER 2/neu; epidermal growth factor receptor; EGFR; multispecific protein;

```

```

DE XX Tetanus toxoid protein derived T-cell activating epitope P2.
KW
KW KW Surface layer protein; S-layer secretion signal; antibiotic; vaccine;
KW recombinant fusion protein cleavage; enzyme; protein polymer; foodstuff;
KW KW antibacterial enzyme; surface glycoprotein; T-cell activating epitope;
KW KW P2; tetanus toxoid; IPNV; Infectious pancreatic necrosis virus.
XX
XX OS Clostridium tetani.
XX
XX PN WO200004170-A1.
XX
XX PD 27-JAN-2000.
XX
XX PF 14-JUL-1999; 99WO-CA000637.
XX
XX PR 14-JUL-1998; 98CA-2237704.
XX
XX PA (UYBR-) UNIV BRITISH COLUMBIA.
XX
XX PI Smit J;
XX
XX DR WPI; 2000-182434/16.
XX
XX PT Cleavage of Caulobacter produced recombinant fusion proteins useful for
XX PT producing vaccine peptides -
XX
XX PS Example 2; Page 16; 33pp; English.
XX
XX CC The patent discloses a method for cleaving a recombinant fusion protein
XX CC which is produced by Caulobacter and consists of Caulobacter surface
XX CC layer (S-layer) protein (containing the C-terminal secretion signal) and
XX CC a target protein heterologous to Caulobacter. The cleavage of target
XX CC protein from the S-layer protein is carried out under mild acid
XX CC conditions so that cleavage occurs at aspartate-proline dipeptide site
XX CC without solubilising the protein. The cleavage is accomplished while the
XX CC fusion protein is in an insoluble aggregate form which facilitates
XX CC purification of the protein. The method is useful for producing pure
XX CC proteins including recombinant human and animal therapeutic antibiotic
XX CC and vaccine peptides, enzymes, protein polymers, and antibacterial
XX CC enzymes for foodstuffs.
XX CC The present sequence is a T-cell activating epitope P2 derived
XX CC from tetanus toxoid protein. This sequence was fused to a DNA encoding
XX CC a fragment of Infectious pancreatic necrosis virus surface glycoprotein
XX CC which is a vaccine candidate. This chimeric protein was in turn fused to
XX CC DNA encoding C. crescentus S-layer secretion signal (corresponds to the
XX CC C-terminal portion of the S-layer protein from amino acid 690 onwards and
XX CC contains native Asp-Pro site) for construction of a recombinant
XX CC fusion construct which is expressed in Caulobacter and then cleaved
XX CC to recover the vaccine candidate protein.
XX
XX SQ Sequence 15 AA;
XX
XX Query Match 100.0%; Score 74; DB 21; Length 15;
XX Best Local Similarity 100.0%; Pred. No. 2.3e-07;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 QYIKANSKFIGITEL 15
XX | | | | | | | | | |
XX Db 1 qyikanskfigitel 15
XX
XX RESULT 15
XX AAB85451
XX ID AAB85451 standard; peptide; 15 AA.
XX
XX AC AAB85451;
XX
XX DT 25-SEP-2001 (first entry)
XX
XX DE Wild-type TT830 (tetanus toxin) epitope.
XX
XX KW HER 2/neu; epidermal growth factor receptor; EGFR; multispecific protein;

```

KW Fc receptor; FcR; tumor cell; breast; cancer; sarcoma; carcinoma; HIV;  
KW pathogenic; Toxoplasma gondii; candidiasis; systemic lupus; cytostatic;  
KW immune thrombocytopenia purpura; immunosuppressive; antiviral;  
XX antifungal; antiprotozoal; TT830; tetanus toxin.

OS Clostridium tetani.

XX US6270765-B1.

PN 07-AUG-2001.

XX 06-NOV-1998; 98US-0188082.

XX 07-JUN-1996; 96US-0661052.

PR 07-JUN-1995; 95US-0484172.

XX (MEDA-) MEDAREX INC.

XX Deco YM, Goldstein J, Graziano R, Somasundaram C;

XX WPI; 2001-475189/51.

DR N-PSDB; AAH23378.

XX Inducing killing of tumor cells which expresses HER 2/neu or epidermal  
PT growth factor receptor (EGFR) by contacting the cell with multispecific  
PT proteins comprising an anti-Fc receptor, -Her 2/neu or -EGFR antibody,  
PT useful for treating cancer -

XX Example 7; Fig 24; 57pp; English.

XX The invention relates to a new method for inducing killing of a tumor  
CC cell which expresses HER 2/neu or epidermal growth factor receptor  
CC (EGFR). The method comprises contacting the tumor cell with a  
CC multispecific protein comprising a component, preferably an antibody,  
CC which binds to an Fc receptor (FcR), Her 2/neu or EGFR. The method is  
CC useful for inducing killing of a tumor cell from breast cancer, sarcoma,  
CC carcinoma, or ovarian cancer. Specific multispecific proteins can also be  
CC administered to a subject to treat or prevent other diseases or  
CC conditions, including pathogenic infections (e.g., viral (such as HIV)),  
CC protozoan infections (such as Toxoplasma gondii), fungal infections  
CC (such as candidiasis), and an autoimmunity (e.g. immune thrombocytopenia  
CC purpura and systemic lupus). The present sequence represents a wild-type  
CC tetanus toxin TT830 epitope.

XX Sequence 15 AA;

Query Match 100.0%; Score 74; DB 22; Length 15;  
Best Local Similarity 100.0%; Pred. No. 2.3e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QYIRANSKFIGITEL 15  
| | | | | | | | | | | | | | |  
Db 1 qyikanskfigitel 15

Search completed: January 29, 2002, 10:59:20  
Job time: 2244 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 10:59:58 ; Search time 32.24 Seconds  
(without alignments)  
10.470 Million cell updates/sec

Title: us-09-763-397A-24  
Perfect score: 74  
Sequence: 1 QYIKANSKFIGITEL 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles!.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	74	100.0	15	2	US-08-319-704-10
2	74	100.0	15	2	US-08-661-052-6
3	74	100.0	15	2	US-08-460-502-7
4	74	100.0	15	4	US-09-046-373-2
5	74	100.0	15	4	US-09-188-082-6
6	74	100.0	15	5	PCT-US93-11703-69
7	74	100.0	16	4	US-09-248-588-55
8	74	100.0	17	1	US-08-446-692-4
9	74	100.0	17	2	US-08-488-351A-4
10	74	100.0	17	3	US-09-100-409A-40
11	74	100.0	17	5	PCT-US95-08596-23
12	74	100.0	17	5	PCT-US95-13841-7
13	74	100.0	27	1	US-08-446-692-13
14	74	100.0	27	2	US-08-488-351A-13
15	74	100.0	31	5	PCT-US93-11703-63
16	74	100.0	37	1	US-08-446-692-57
17	74	100.0	37	1	US-08-446-692-63
18	74	100.0	37	2	US-08-488-351A-57
19	74	100.0	37	2	US-08-488-351A-63
20	74	100.0	47	1	US-08-446-692-35
21	74	100.0	47	2	US-08-488-351A-35
22	70	94.6	14	1	US-08-186-266-5
23	70	94.6	14	1	US-08-305-871A-5
24	70	94.6	14	1	US-08-465-167A-18
25	70	94.6	14	2	US-08-817-933A-9
26	70	94.6	14	5	PCT-US92-07218-15
27	70	94.6	14	5	PCT-US92-07218-30

28	70	94.6	14	5	PCT-US95-02121-95	Sequence 95, Appl
29	70	94.6	24	5	PCT-US92-07218-25	Sequence 25, Appl
30	70	94.6	24	5	PCT-US92-07218-31	Sequence 31, Appl
31	70	94.6	24	5	PCT-US95-02121-110	Sequence 110, App
32	70	94.6	27	5	PCT-US92-07218-26	Sequence 26, Appl
33	70	94.6	27	5	PCT-US92-07218-27	Sequence 27, Appl
34	70	94.6	27	5	PCT-US92-07218-28	Sequence 28, Appl
35	70	94.6	27	5	PCT-US92-07218-32	Sequence 32, Appl
36	70	94.6	27	5	PCT-US95-02121-111	Sequence 111, App
37	70	94.6	27	5	PCT-US95-02121-112	Sequence 112, App
38	70	94.6	29	3	US-09-075-257A-13	Sequence 13, Appl
39	70	94.6	29	3	US-09-075-257A-14	Sequence 14, Appl
40	70	94.6	29	4	US-09-534-639-13	Sequence 13, Appl
41	70	94.6	29	4	US-09-534-639-14	Sequence 14, Appl
42	70	94.6	30	5	PCT-US92-07218-29	Sequence 29, Appl
43	70	94.6	32	1	US-08-186-266-9	Sequence 9, Appl
44	70	94.6	50	4	US-09-171-969-7	Sequence 7, Appl
45	69	93.2	15	2	US-08-661-052-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1  
US-08-319-704-10  
; Sequence 10, Application US/08319704  
; Patent No. 5814617  
; GENERAL INFORMATION:  
; APPLICANT: Hoffman, Stephen L.  
; APPLICANT: Charoenvit, Yupin  
; APPLICANT: Hedstrom, Richard C.  
; APPLICANT: Doolan, Denise L.  
; TITLE OF INVENTION: Protective 17 kDa Malaria Hepatic and  
; TITLE OF INVENTION: Erythrocytic Stage Immunogen and Gene  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Naval Medical R & D Command  
; STREET: Bldg 1, T-12, 8901 Wisconsin Avenue  
; CITY: Bethesda  
; STATE: Maryland  
; COUNTRY: U.S.A  
; ZIP: 20889-5606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/319,704  
; FILING DATE: 07-OCT-1994  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: A. David Spevack  
; REGISTRATION NUMBER: 24,743  
; REFERENCE/DOCKET NUMBER: 75,206  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 295-6759  
; TELEFAX: (301) 295-1022  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 15 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-319-704-10

Query Match 100.0%; Score 74; DB 2; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.2e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 QYIKANSKFIGITEL 15  
|||||

Db 1 QYIKANSKFIGITEL 15

RESULT 2

US-08-661-052-6

; Sequence 6, Application US/08661052

; Patent No. 5837243

; GENERAL INFORMATION:

; APPLICANT: Yashwant M. Deo

; APPLICANT: Joel Goldstein

; APPLICANT: Robert Graziano

; APPLICANT: Chezman Somasundaram

; TITLE OF INVENTION: THERAPEUTIC COMPOUNDS COMPRISED

; TITLE OF INVENTION: OF ANTI-PC RECEPTOR ANTIBODIES

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: LAHIVE & COCKFIELD

; STREET: 60 State Street, Suite 510

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02109-1875

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/661,052

; FILING DATE:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/484,172

; FILING DATE: 07-JUNE-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Arnold, Beth E.

; REGISTRATION NUMBER: 35,430

; REFERENCE/DOCKET NUMBER: MXI-043CP

; TELEPHONE: (617)227-7400

; TELEFAX: (617)227-5941

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 15 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; FRAGMENT TYPE: internal

US-08-661-052-6

Query Match 100.0%; Score 74; DB 2; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.2e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15

Db 1 QYIKANSKFIGITEL 15

RESULT 3

US-08-460-502-7

; Sequence 7, Application US/08460502

; Patent No. 5843464

; GENERAL INFORMATION:

; APPLICANT: Bakaletz, Lauren O.

; APPLICANT: Kaumaya, Parvin T.

; TITLE OF INVENTION: Synthetic Chimeric Flmbrin Peptides

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Calfee, Halter and Griswold

; STREET: 800 Superior Avenue

; CITY: Cleveland

; STATE: Ohio

US-08-460-502-7

Query Match 100.0%; Score 74; DB 4; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.2e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15

Db 1 QYIKANSKFIGITEL 15

RESULT 4

US-09-046-373-2

; Sequence 2, Application US/09046373

; Patent No. 6235714

; GENERAL INFORMATION:

; APPLICANT: Sudhir Paul

; APPLICANT: Larry J. Smith

; APPLICANT: Gennady Gololobov

; TITLE OF INVENTION: Methods for Identifying Inducers and

; TITLE OF INVENTION: Inhibitors of Catalytic Antibodies, Compositions and Their

; TITLE OF INVENTION: Use

; FILE REFERENCE: UNMC 63123

; CURRENT APPLICATION NUMBER: US/09/046,373

; CURRENT FILING DATE: 1998-03-23

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: FastSEQ for Windows version 3.0

; SEQ ID NO 2

; LENGTH: 15

; TYPE: PRT

; ORGANISM: Clostridium tetani

US-09-046-373-2

Query Match 100.0%; Score 74; DB 4; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.2e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15

Db 1 QYIKANSKFIGITEL 15

RESULT 5

US-09-188-082-6

; Sequence 6, Application US/09188082

; Patent No. 6270765

US-09-188-082-6

Query Match 100.0%; Score 74; DB 4; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.2e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15

Db 1 QYIKANSKFIGITEL 15

US-09-188-082-6

Query Match 100.0%; Score 74; DB 4; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.2e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15

Db 1 QYIKANSKFIGITEL 15

US-09-188-082-6

Query Match 100.0%; Score 74; DB 4; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.2e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15

Db 1 QYIKANSKFIGITEL 15

US-09-188-082-6

Query Match 100.0%; Score 74; DB 4; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.2e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15

Db 1 QYIKANSKFIGITEL 15

US-09-188-082-6

Query Match 100.0%; Score 74; DB 4; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.2e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15

Db 1 QYIKANSKFIGITEL 15

US-09-188-082-6

Query Match 100.0%; Score 74; DB 4; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.2e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15

Db 1 QYIKANSKFIGITEL 15

US-09-188-082-6

Query Match 100.0%; Score 74; DB 4; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.2e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15

Db 1 QYIKANSKFIGITEL 15

US-09-188-082-6

Query Match 100.0%; Score 74; DB 4; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.2e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15

Db 1 QYIKANSKFIGITEL 15

US-09-188-082-6

Query Match 100.0%; Score 74; DB 4; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.2e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15

Db 1 QYIKANSKFIGITEL 15

US-09-188-082-6

Query Match 100.0%; Score 74; DB 4; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.2e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15

Db 1 QYIKANSKFIGITEL 15

US-09-188-082-6

Query Match 100.0%; Score 74; DB 4; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.2e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15

Db 1 QYIKANSKFIGITEL 15

US-09-188-082-6

Query Match 100.0%; Score 74; DB 4; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.2e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15

Db 1 QYIKANSKFIGITEL 15

US-09-188-082-6

Query Match 100.0%; Score 74; DB 4; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.2e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15

Db 1 QYIKANSKFIGITEL 15

US-09-188-082-6

Query Match 100.0%; Score 74; DB 4; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.2e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15

Db 1 QYIKANSKFIGITEL 15

US-09-188-082-6

Query Match 100.0%; Score 74; DB 4; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.2e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15

Db 1 QYIKANSKFIGITEL 15

US-09-188-082-6

Query Match 100.0%; Score 74; DB 4; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.2e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15

Db 1 QYIKANSKFIGITEL 15

US-09-188-082-6

Query Match 100.0%; Score 74; DB 4; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.2e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15

Db 1 QYIKANSKFIGITEL 15

US-09-188-082-6

Query Match 100.0%; Score 74; DB 4; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.2e-07;

Matches 15; Conservative 0;

GENERAL INFORMATION:  
APPLICANT: Yashwant M. Deo  
APPLICANT: Joel Goldstein  
APPLICANT: Robert Graziano  
APPLICANT: Chezian Somsundaram  
TITLE OF INVENTION: THERAPEUTIC COMPOUNDS COMPRISED  
OF ANTI-FC RECEPTOR ANTIBODIES  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street, Suite 510  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/188,082  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/661,052  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold, Beth E.  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: MXI-043CP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-09-188-082-6

Query Match 100.0%; Score 74; DB 4; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.2e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15  
DB 1 QYIKANSKFIGITEL 15

RESULT 6  
PCT-US93-11703-69  
GENERAL INFORMATION:  
APPLICANT: Chiron Mimotopes Pty. Ltd.  
TITLE OF INVENTION: T-Cell Epitopes  
NUMBER OF SEQUENCES: 75  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Grant D. Green  
STREET: 4560 Horton St.  
CITY: Emeryville  
STATE: CA  
COUNTRY: USA  
ZIP: 94608  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/11703

FILING DATE: 28-DEC-1993  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/984,852  
FILING DATE: 02-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Green, Grant D.  
REGISTRATION NUMBER: 31,259  
REFERENCE/DOCKET NUMBER: 0222.101  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 510-601-2706  
TELEFAX: 510-655-3542  
INFORMATION FOR SEQ ID NO: 69:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US93-11703-69

Query Match 100.0%; Score 74; DB 5; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.2e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15  
DB 1 QYIKANSKFIGITEL 15

RESULT 7  
US-09-248-588-55  
Sequence 55, Application US/09248588  
Patent No. 6231864  
GENERAL INFORMATION:  
APPLICANT: Birkett, Ashley J.  
TITLE OF INVENTION: Strategically Modified Hepatitis B Core Proteins and  
their Derivatives  
FILE REFERENCE: SYN-101 4564/69529  
CURRENT APPLICATION NUMBER: US/09/248,588  
CURRENT FILING DATE: 1999-02-11  
EARLIER APPLICATION NUMBER: 60/074537  
EARLIER FILING DATE: 1998-02-12  
NUMBER OF SEQ ID NOS: 113  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 55  
LENGTH: 16  
TYPE: PRT  
ORGANISM: Clostridium tetani  
PUBLICATION INFORMATION:  
JOURNAL: Vaccine  
VOLUME: 15  
ISSUE: 4  
PAGES: 377-  
DATE: 1997  
US-09-248-588-55

Query Match 100.0%; Score 74; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 1.3e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15  
DB 1 QYIKANSKFIGITEL 15

RESULT 8  
US-08-446-692-4  
Sequence 4, Application US/08446692  
Patent No. 5759551  
GENERAL INFORMATION:

APPLICANT: Ladd, Anna  
APPLICANT: Wang, Chang Yi  
APPLICANT: Zamb, Timothy  
TITLE OF INVENTION: Immunogenic LHRH peptide constructs  
TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines  
NUMBER OF SEQUENCES: 114  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Maria C.H. Lin  
STREET: 345 Park Avenue  
CITY: New York  
STATE: NY  
COUNTRY: US  
ZIP: 10154-0053  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/446,692  
FILING DATE: 7-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maria C.H. Lin  
REGISTRATION NUMBER: 29,323  
REFERENCE/DOCKET NUMBER: 1151-4146 US2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)415-8745  
TELEFAX: (516)751-6849  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-446-692-4

Query Match 100.0%; Score 74; DB 1; Length 17;  
Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QYIKANSKFIGITEL 15  
Db 3 QYIKANSKFIGITEL 17  
RESULT 9  
US-08-488-351A-4  
Sequence 4, Application US/08488351A  
Patent No. 5843446  
GENERAL INFORMATION:  
APPLICANT: Ladd, Anna  
APPLICANT: Wang, Chang Yi  
APPLICANT: Zamb, Timothy  
TITLE OF INVENTION: Immunogenic LHRH peptide constructs  
TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines  
NUMBER OF SEQUENCES: 114  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Maria C.H. Lin  
STREET: 345 Park Avenue  
CITY: New York  
STATE: NY  
COUNTRY: US  
ZIP: 10154-0053  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/488,351A  
FILING DATE: 7-JUN-1995

CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/446,692  
FILING DATE: 7-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/229,275  
FILING DATE: 14-APR-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/057,166  
FILING DATE: 27-APR-1992  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maria C.H. Lin  
REGISTRATION NUMBER: 29,323  
REFERENCE/DOCKET NUMBER: 1151-4146 US2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)415-8745  
TELEFAX: (516)751-6849  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-488-351A-4

Query Match 100.0%; Score 74; DB 2; Length 17;  
Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QYIKANSKFIGITEL 15  
Db 3 QYIKANSKFIGITEL 17

RESULT 10  
US-09-100-409A-40  
Sequence 40, Application US/09100409A  
Patent No. 6090388  
GENERAL INFORMATION:  
APPLICANT: Wang, Chang Yi  
TITLE OF INVENTION: PEPTIDE COMPOSITION FOR  
PREVENTION AND TREATMENT OF HIV INFECTION AND  
IMMUNE DISORDERS  
TITLE OF INVENTION: IMMUNE DISORDERS  
NUMBER OF SEQUENCES: 64  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 Park Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10154-0054  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version  
#1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/100,409A  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME:  
REGISTRATION NUMBER:  
REFERENCE/DOCKET NUMBER: 1151-4154  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-758-4800  
TELEFAX: 212-751-6849  
INFORMATION FOR SEQ ID NO: 40:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 17 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-100-409A-40

Query Match 100.0%; Score 74; DB 3; Length 17;  
Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QYIKANSKFIGITEL 15  
| | | | | | | | | | | | | | | | |  
Db 3 QYIKANSKFIGITEL 17

RESULT 11  
PCT-US95-08596-23  
; Sequence 23, Application PC/TUS9508596  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: Proinsulin Peptide Compounds for Detecting  
; and treating Type I Diabetes  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 60 State Street, suite 510  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109-1875

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/08596  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/272,220  
; FILING DATE: 08-JULY-1994

; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: DeConti, Giulio A., Jr.  
; REGISTRATION NUMBER: 31,503  
; REFERENCE/DOCKET NUMBER: DCI-092PC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)227-7400  
; TELEFAX: (617)227-5941  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 17 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: internal  
PCT-US95-08596-23

Query Match 100.0%; Score 74; DB 5; Length 17;  
Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QYIKANSKFIGITEL 15  
| | | | | | | | | | | | | | | | |  
Db 3 QYIKANSKFIGITEL 17

RESULT 12  
PCT-US95-13841-7

; Sequence 7, Application PC/TUS9513841  
; GENERAL INFORMATION:  
; APPLICANT: United Biomedical Inc; Walfield, Alan M.;  
; APPLICANT: Wang, Chang Yi  
; TITLE OF INVENTION: Synthetic IgE Membrane Anchor  
; TITLE OF INVENTION: Peptide Immunogens for the Treatment of Allergy  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Maria C.H. Lin  
; STREET: 345 Park Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WordPerfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/13841  
; FILING DATE: 25-OCT-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/328,519  
; FILING DATE: 25-OCT-1994  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lin, Maria C.H.  
; REGISTRATION NUMBER: 29,323  
; REFERENCE/DOCKET NUMBER: 1151-4117  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-758-4800  
; TELEFAX: 212-751-6849  
; TELEX: 421792  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 17 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
PCT-US95-13841-7

Query Match 100.0%; Score 74; DB 5; Length 17;  
Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QYIKANSKFIGITEL 15  
| | | | | | | | | | | | | | | | |  
Db 3 QYIKANSKFIGITEL 17

RESULT 13  
US-08-446-692-13  
; Sequence 13, Application US/08446692  
; Patent No. 5759551  
; GENERAL INFORMATION:  
; APPLICANT: Ladd, Anna  
; APPLICANT: Wang, Chang Yi  
; APPLICANT: Zamb, Timothy  
; TITLE OF INVENTION: Immunogenic LHRH peptide constructs  
; and synthetic universal immune stimulators for vaccines  
; NUMBER OF SEQUENCES: 114  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Maria C.H. Lin  
; STREET: 345 Park Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: US  
; ZIP: 10154-0053  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

Qy	1	QYIKANSKFIGITEL	15
Db	10	QYIKANSKFIGITEL	24



Search completed: January 29, 2002, 10:59:58  
Job time: 2144 sec

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 11:00:40 ; Search time 34.94 Seconds  
(without alignments)  
32.702 Million cell updates/sec

Title: US-09-763-397A-24

Perfect score: 74

Sequence: 1 QYIKANSKFIGITEL 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_68.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	74	100.0	1315	1 BRCLTN	tentoxylisin (EC 3
2	44.5	60.1	244	2 S29982	class II histocomp
3	44	59.5	66	2 S31029	gene 84 protein -
4	43	58.1	180	2 G86826	diamine N-acetyltr
5	43	58.1	899	2 T42976	hypothetical prote
6	42.5	57.4	1060	2 S06286	major merizoite su
7	42.5	57.4	1086	2 S16752	major merizoite su
8	42.5	57.4	1701	2 A54498	major merizoite su
9	42.5	57.4	1701	2 A26868	major merizoite su
10	42.5	57.4	1726	1 SAZQGM	major merizoite su
11	42.5	57.4	1726	2 A45948	major merizoite su
12	42	56.8	1333	2 S38635	blastopia polyprot
13	41	55.4	123	2 G48677	Ig heavy chain v-D
14	41	55.4	899	2 G36812	hypothetical prote
15	40.5	54.7	245	2 S29980	class II histocomp
16	40	54.1	79	2 D85794	hypothetical prote
17	40	54.1	194	2 G64026	lacyl-carrier-prot
18	40	54.1	601	1 A55485	oligopeptidase (EC
19	40	54.1	601	2 G86840	oligoendopeptidase
20	40	54.1	644	2 S46746	hypothetical prote
21	39	52.7	102	2 PH1491	Ig heavy chain v r
22	39	52.7	119	2 PH1516	Ig heavy chain v r
23	39	52.7	119	2 PH1518	Ig heavy chain v r
24	39	52.7	119	2 PH1519	Ig heavy chain v r
25	39	52.7	123	2 F48677	Ig heavy chain v-D
26	39	52.7	135	2 PH1494	Ig heavy chain v r
27	39	52.7	140	2 PH1488	Ig heavy chain v r
28	39	52.7	213	1 KTYMC	adenylate kinase (
29	39	52.7	326	2 B71808	type II restrictio

#### RESULT 1

BRCLTN

tentoxylisin (EC 3.4.24.68) precursor - Clostridium tetani

N:Alternate names: tetanus neurotoxin

C:Species: Clostridium tetani

C:Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text change 18-Jun-1999

C:Accession: A25689; A25757; A25194; B25194; A60759; S69348; S09364

R:Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels, J.; Weller, U.; Hudel, EMBO J. 5, 2495-2502, 1986

A:Title: Tetanus toxin: Primary structure, expression in E. coli, and homology with b

A:Reference number: A25689; MUID:87053814

A:Accession: A25689

A:Molecule type: DNA

A:Residues: 1-1315 <EIS>

A:Cross-references: GB:X04436; NID:g40769; PIDN:CAA28033.1; PID:g40770

R:Fairweather, N.F.; Lyness, V.A.

Nucleic Acids Res. 14, 7809-7812, 1986

A:Title: The complete nucleotide sequence of tetanus toxin.

A:Reference number: A25757; MUID:87040747

A:Accession: A25757

A:Molecule type: DNA

A:Residues: 1-1315 <FAI>

A:Cross-references: GB:X06214; NID:g40773; PIDN:CAA29564.1; PID:g40774

A:Experimental source: strain CN3911

R:Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.

J. Bacteriol. 165, 21-27, 1986

A:Title: Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C 1

A:Reference number: A25194; MUID:86085672

A:Accession: A25194

A:Molecule type: DNA

A:Residues: 743-1315 <FA2>

A:Cross-references: GB:M12739; NID:g144920; PIDN:AAA23282.1; PID:g144921

A:Accession: B25194

A:Molecule type: protein

R:Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.

Infect. Immun. 57, 3588-3593, 1989

A:Title: Isolation, purification, and characterization of fragment B, the NH-2-termin

A:Reference number: A60759; MUID:90035436

A:Accession: A60759

A:Molecule type: protein

A:Residues: 461-475 <MAT>

R:Demotz, G.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.

J. Immunol. 142, 394-402, 1989

A:Title: Delineation of several DR-restricted tetanus toxin T cell epitopes.

A:Reference number: J50098; MUID:89093918

A:Contents: annotation; epitope region

R:Schlavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta,

Nature 359, 832-835, 1992

A:Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteo

A:Reference number: S27125; MUID:93063293

A:Contents: .annotation

R:de Fillippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.; Montecucco, C.  
Eur. J. Biochem. 229, 61-69, 1995  
A:Title: Structural studies on the zinc-endopeptidase light chain of tetanus neurotoxin.  
A:Reference number: S69348; MUID:9526288  
A:Accession: S69348  
A:Molecule type: protein  
A:Residues: 2-31 <DEF>  
C:Comment: The source of this protein was an extrachromosomal plasmid.  
C:Comment: The precursor is cleaved by endogenous proteinase activity to form light (fragment B) and heavy (fragment A) chains. The amino end of the heavy chain (fragment B) forms ion channels in a lipid bilayer. Fragment C binds to ganglionic presynaptic neurons. It inhibits neurotransmitter release by proteolytic cleavage of synaptobrevin.  
C:Function:  
A:Description: blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide bond in synaptobrevin.  
C:Superfamily: tetanus toxin  
C:Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc  
F:2-457/Product: tentoxylisin light chain (fragment A) #status predicted <TTL>  
F:461-1315/Product: tentoxylisin heavy chain (fragment B.C) #status experimental <TTH>  
F:461-864/Domain: channel forming (fragment B) #status predicted <TXB>  
F:865-1315/Domain: ganglioside binding (fragment C) #status predicted <TXC>  
F:233,237/Binding site: zinc (His) #status predicted  
F:234/Active site: Glu #status predicted

Query Match 100.0%; Score 74; DB 1; Length 1315;  
Best Local Similarity 100.0%; Pred. No. 2.8e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15  
|||||  
Db 830 QYIKANSKFIGITEL 844

RESULT 2  
S29982  
class II histocompatibility antigen - Atlantic salmon  
C:Species: Salmo salar (Atlantic salmon)  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 20-Sep-1999  
C:Accession: S29982  
R:Hardvik, I.  
submitted to the EMBL Data Library, October 1992  
A:Reference number: S29980  
A:Accession: S29982  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-244 <HOR>  
A:Cross-references: EMBL:X70166; NID:g64369; PID:g64370  
C:Superfamily: class II histocompatibility antigen; immunoglobulin homology

Query Match 60.1%; Score 44.5; DB 2; Length 244;  
Best Local Similarity 55.6%; Pred. No. 1.4;  
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 1 QYIKANS---KFIGITEL 15  
:||: || |||  
Db 51 EYIRFNSTVGKFGVTEV 68

RESULT 3  
S31029  
gene 84 protein - Mycobacterium phage L5  
C:Species: Mycobacterium phage L5  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 08-Oct-1999  
C:Accession: S31029  
R:Donnelly-Wu, M.K.; Jacobs Jr., W.R.; Hatfull, G.F.  
Mol. Microbiol. 7, 407-417, 1993  
A:Title: Superinfection immunity of mycobacteriophage L5: applications for genetic transduction.  
A:Reference number: S30949; MUID:93211283  
A:Accession: S31029  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA

A:Residues: 1-66 <DON>  
A:Cross-references: EMBL:Z18946; NID:g15859; PIDN:CAA79460.1; PID:e59702; PID:g579152  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1992  
C:Genetics: 84  
A:Gene: 84  
A:Start codon: GTG

Query Match 59.5%; Score 44; DB 2; Length 66;  
Best Local Similarity 57.1%; Pred. No. 0.45;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 YIKANSKFIGITEL 15  
||| |||  
Db 50 YIKRNGKFGVTWEV 63

RESULT 4  
G86826  
diamine N-acetyltransferase (EC 2.3.1.57) [imported] - Lactococcus lactis subsp. lact  
N:Alternate names: spermidine acetyltransferase  
C:Species: Lactococcus lactis subsp. lactis  
C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 04-Apr-2001  
C:Accession: G86826  
R:Botolin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Eh  
Genome Res. in press, 2001  
A:Title: The complete genome sequence of the lactic acid bacterium.  
A:Reference number: A86625  
A:Accession: G86826  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-180 <STO>  
A:Cross-references: GB:AE005176; NID:gl2724622; PIDN:AAK05713.1; GSPDB:GN00146  
A:Experimental source: strain IL1403  
C:Genetics:  
C:Gene: YGF  
C:Keywords: acyltransferase; coenzyme A

Query Match 58.1%; Score 43; DB 2; Length 180;  
Best Local Similarity 69.2%; Pred. No. 1.9;  
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 IKANSKFIGITEL 15  
|-|| |||||  
Db 65 IEANDFIGIVEL 77

RESULT 5  
T42976  
hypothetical protein 63 - ateline herpesvirus 3 (strain 73)  
C:Species: ateline herpesvirus 3  
A:Variety: Strain 73  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jan-2000  
C:Accession: T42976  
R:Albrecht, J.C.; Fleckenstein, B.  
submitted to the EMBL Data Library, August 1998  
A:Description: Primary structure of the herpesvirus ateles genome.  
A:Reference number: Z22274  
A:Accession: T42976  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-899 <ALB>  
A:Cross-references: EMBL:AF083424; PIDN:AAC95587.1  
A:Experimental source: strain 73

Query Match 58.1%; Score 43; DB 2; Length 899;  
Best Local Similarity 64.3%; Pred. No. 9.9;  
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITE 14  
||| :||| |||



C;Accession: S29980  
R;Hordvik, I.  
submitted to the EMBL Data Library, October 1992  
A;Reference number: S29980  
A;Accession: S29980

A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-245 <HOR>  
A:Cross-references: EMBL:X70167; NID:g64371; PID:g64372  
C:Superfamily: class II histocompatibility antigen; immunoglobulin homology

Query Match 54.7%; Score 40.5; DB 2; Length 245;  
Best Local Similarity 44.4%; Pred. No. 7.6;  
Matches 8; Conservative 5; Mismatches 2; Indels 3; Gaps 1;

Qy 1 QYIKANS---KFIGITEL 15  
:|::||| |::|  
Db 53 EYVRFNSTVGKXVGYTEL 70

Search completed: January 29, 2002, 11:00:40  
Job time: 2030 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 11:16:19 ; Search time 20.36 Seconds  
(without alignments)  
27.012 Million cell updates/sec

Title: US-09-763-397A-24

Perfect score: 74

Sequence: 1 QYIKANSKFIGITEL 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	74	100.0	1314	1	TETX_CLOTE
2	44	59.5	66	1	VG84_CLOTE
3	42.5	57.4	1682	1	MSPL_PLAF3
4	42.5	57.4	1701	1	MSPL_PLAFM
5	42.5	57.4	1701	1	MSPL_PLAFM
6	42.5	57.4	1726	1	MSPL_PLAF3
7	42.5	57.4	1726	1	MSPL_PLAFM
8	41	55.4	204	1	PYRC_SERMA
9	41	55.4	899	1	V120_HSVSA
10	40	54.1	194	1	ACPD_HAEIN
11	40	54.1	601	1	PEFL_LACLC
12	40	54.1	644	1	YHJ9_YEAST
13	38	52.7	213	1	KAD_MYCCA
14	38	51.4	256	1	YD83_METJA
15	38	51.4	287	1	TRUB_AQUAE
16	38	51.4	572	1	HEWA_PI3HT
17	37	50.0	191	1	Y096_HAEIN
18	37	50.0	445	1	GNT1_HUMAN
19	37	50.0	447	1	GNT1_MOUSE
20	37	50.0	447	1	GNT1_RABIT
21	37	50.0	447	1	GNT1_RAT
22	37	50.0	490	1	Y032_BORBU
23	37	50.0	510	1	G6PD_ASPNG
24	37	50.0	511	1	G6PD_EMENI
25	37	50.0	548	1	YDD2_SCHPO
26	36	48.6	169	1	Y358_BUCAI
27	36	48.6	258	1	MIP_CHLPN
28	36	48.6	296	1	YD01_CLOB
29	36	48.6	333	1	YD08_XENLA
30	36	48.6	451	1	MURD_BACSU
31	36	48.6	461	1	NIFN_RHOCA
32	36	48.6	495	1	G6PD_PICJA
33	36	48.6	630	1	YND1_YEAST

## RESULT 1

ID	TETX_CLOTE	STANDARD;	PRT;	1314 AA.
AC	P04958;			
DT	13-AUG-1987 (Rel. 05, Created)			
DT	13-AUG-1987 (Rel. 05, Last sequence update)			
DE	TETANUS TOXIN PRECURSOR (EC 3.4.24.68) (TENTOXLYSIN).			
OS	Clostridium tetani.			
OG	Plasmid.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;			
OC	Clostridium.			
OX	NCBI_TaxId=1513;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=87053814; PubMed=3536478;			
RA	Eisel U., Jarausch W., Goretzki K., Henschen A., Engels J.,			
RA	Weller U., Hudel M., Habermann E., Niemann H.;			
RT	"Tetanus toxin: primary structure, expression in E. coli, and			
RT	homology with botulinum toxins.";			
RL	EMBO J. 5:2495-2502(1986).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CN3911;			
RX	MEDLINE=87040747; PubMed=3774547;			
RA	Fairweather N.F., Lyness V.A.;			
RT	"The complete nucleotide sequence of tetanus toxin.";			
RL	Nucleic Acids Res. 14:7809-7812(1986).			
RN	[3]			
RP	SEQUENCE OF 742-1314 FROM N.A.			
RX	MEDLINE=86085672; PubMed=3510187;			
RA	Fairweather N.F., Lyness V.A., Pickard D.J., Allen G., Thomson R.O.;			
RT	"Cloning, nucleotide sequencing, and expression of tetanus toxin			
RT	fragment C in Escherichia coli.";			
RL	J. Bacteriol. 165:21-27(1986).			
RN	[4]			
RP	PARTIAL SEQUENCE, AND DISULFIDE BONDS.			
RX	MEDLINE=90201034; PubMed=2108021;			
RA	Kriegstein K., Henschen A., Weller U., Habermann E.;			
RT	"Arrangement of disulfide bridges and positions of sulphhydryl groups			
RT	in tetanus toxin.";			
RL	Eur. J. Biochem. 188:39-45(1990).			
RN	[5]			
RP	PARTIAL SEQUENCE.			
RX	MEDLINE=92037649; PubMed=1935979;			
RA	Kriegstein K.G., Henschen A.H., Weller U., Habermann E.;			
RT	"Limited proteolysis of tetanus toxin. Relation to activity and			
RT	identification of cleavage sites.";			
RL	Eur. J. Biochem. 202:41-51(1991).			
RN	[6]			
RP	IDENTIFICATION AS ZINC-PROTEASE.			
RX	MEDLINE=93010948; PubMed=1396558;			
RA	Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,			
RA	Montecucco C.;			
RT	"Tetanus toxin is a zinc protein and its inhibition of			
RT	neurotransmitter release and protease activity depend on zinc.";			

## ALIGNMENTS

34	36	48.6	774	1	RRP3_INCBE
35	36	48.6	774	1	RRP3_INCJJ
36	36	48.6	1630	1	MSPL_PLAFK
37	36	48.6	1639	1	MSPL_PLAFW
38	35	47.3	176	1	NU6C_SPIOL
39	35	47.3	261	1	CABV_CHICK
40	35	47.3	294	1	CDD_ECOLI
41	35	47.3	321	1	YDG7_SCHPO
42	35	47.3	329	1	DHOA_EMENI
43	35	47.3	353	1	41KD_LACHE
44	35	47.3	402	1	VGLD_HSVEA
45	35	47.3	431	1	Y4OP_RHISN

P21770	Influenza c
P13877	Influenza c
P04932	plasmodium
P04933	plasmodium
Q9m318	spinacia ol
P04354	gallus gall
P13652	escherichia
Q10494	schizosacch.
P25415	mericicella
P17212	lactobacilli
P24872	equine hepr
P55601	rhizobium s

```

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QYIKANSKFITIGTEL 15
    |||||
Db 829 QYIKANSKFITIGTEL 843

RESULT 2
VG84_BPML5          STANDARD;          PRT;          66 AA.
AC Q05301;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE GENE 84 PROTEIN (GP84).
GN 84.
OS Mycobacteriophage L5.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.
OX NCBI_TaxID=31757;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93211282; PubMed=8459766;
RA Hatfull G.F., Sarkis G.J.;
RT "DNA sequence, structure and gene expression of mycobacteriophage L5:
  a phage system for mycobacterial genetics.";
RL Mol. Microbiol. 7:395-405(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z18946; CAA79460.1; -.
DR DR
DR PIR; S31029; S31029.
SQ SEQUENCE 66 AA; 7424 MW; 9C7104C7A4FA74A5 CRC64;

Query Match 59.5%; Score 44; DB 1; Length 66;
Best Local Similarity 57.1%; Pred. No. 0.2;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 2 YIKANSKFITIGTEL 15
    ||| | |||
Db 50 YIKRNGKPVGTWEV 63

RESULT 3
MSPL_PLAF3          STANDARD;          PRT;          1682 AA.
ID MSPL_PLAF3
AC P19598; Q25921;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMMSA) (P190).
DE DE
GN MSP-1.
OS Plasmodium falciparum (isolate ro-33 / Ghana).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5834;
RN [1]
RP SEQUENCE OF 1-1061 FROM N.A.
RX MEDLINE=88166657; PubMed=3327688;
RA Certa U., Rotmann D., Matile H., Reber-Liske R.;
RT "A naturally occurring gene encoding the major surface antigen
  precursor p190 of Plasmodium falciparum lacks tripeptide repeats.";
RL EMBO J. 6:4137-4142(1987).
RN [2]
RP SEQUENCE OF 1032-1682 FROM N.A.
RX MEDLINE=95354793; PubMed=7628566;
RA Tolle R., Buijard H., Cooper J.A.;

```

RT \*Plasmodium falciparum: variations within the C-terminal region of  
RL merozoite surface antigen-1.\*;  
CC Exp. Parasitol. 81:47-54(1995).  
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR  
CC (POTENTIAL).  
CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42  
CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF  
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M35727; AAA29715.1; -;  
DR EMBL; Y00087; CAA68280.1; -;  
DR EMBL; Z35326; CAA84555.1; -;  
DR PIR; S06286; S06286.  
DR InterPro: IPR000561; EGF-like.  
DR Pfam: PF00008; EGF; 1.  
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;  
KW Transmembrane; GPI-anchor.  
FT SIGNAL 1 19  
FT CHAIN 20 1682  
FT TRANSMEM 1666 1682  
FT CARBOHYD 233 233  
FT CARBOHYD 462 462  
FT CARBOHYD 528 528  
FT CARBOHYD 599 599  
FT CARBOHYD 785 785  
FT CARBOHYD 881 881  
FT CARBOHYD 901 901  
FT CARBOHYD 947 947  
FT CARBOHYD 1071 1071  
FT CARBOHYD 1178 1178  
FT CARBOHYD 1569 1569  
SQ SEQUENCE 1682 AA; 192462 MW; C82A1E159948CAD6 CRC64;  
  
Query Match 57.4%; Score 42.5; DB 1; Length 1682;  
Best Local Similarity 60.0%; Pred. No. 11;  
Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;  
  
QY 1 QYIKANSKFI-GITE 14  
Db 983 QFVKSNSKVITGLTE 997  
I:::|||||I|I|I|  
  
RESULT 4  
MSPL\_PLAFF STANDARD; PRT; 1701 AA.  
AC P13819;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)  
DE (PMMSA).  
GN MSP-1.  
OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5837;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88142999; PubMed=2449612;  
RA Peterson M.G., Coppel R.L., McIntyre P., Langford C.J., Woodrow G.,  
RA Brown G.V., Anders R.F., Kemp D.J.;  
RT "Variation in the precursor to the major merozoite surface antigens  
RT of Plasmodium falciparum.";  
RL Mol. Biochem. Parasitol. 27:291-302(1988).  
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR

CC (POTENTIAL).  
CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42  
CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF  
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M19143; AAA29653.1; -;  
DR PIR; A54498; A54498.  
DR InterPro: IPR000561; EGF-like.  
DR Pfam: PF00008; EGF; 1.  
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;  
KW Transmembrane; GPI-anchor.  
FT SIGNAL 1 19  
FT CHAIN 20 1701  
FT CARBOHYD 110 110  
FT CARBOHYD 239 239  
FT CARBOHYD 470 470  
FT CARBOHYD 536 536  
FT CARBOHYD 607 607  
FT CARBOHYD 802 802  
FT CARBOHYD 899 899  
FT CARBOHYD 919 919  
FT CARBOHYD 965 965  
FT CARBOHYD 991 991  
FT CARBOHYD 1089 1089  
FT CARBOHYD 1196 1196  
FT CARBOHYD 1588 1588  
SQ SEQUENCE 1701 AA; 193719 MW; 3920B75E73D38552 CRC64;  
  
Query Match 57.4%; Score 42.5; DB 1; Length 1701;  
Best Local Similarity 60.0%; Pred. No. 11;  
Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;  
  
QY 1 QYIKANSKFI-GITE 14  
Db 1001 QFVKSNSKVITGLTE 1015  
I:::|||||I|I|I|  
  
RESULT 5  
MSPL\_PLAFF STANDARD; PRT; 1701 AA.  
AC P08569;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)  
DE (PMMSA) (P190).  
GN MSP-1.  
OS Plasmodium falciparum (isolate mad20 / Papua New Guinea).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=70153;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88011243; PubMed=3079521;  
RA Tanabe K., Mackay M., Goman M., Scaife J.G.;  
RT "Allelic dimorphism in a surface antigen gene of the malaria parasite  
RT Plasmodium falciparum.";  
RL J. Mol. Biol. 195:273-287(1987).  
RN [2]  
RP REVISIONS TO 1403; 1569 AND 1629.  
RA Tanabe K.;  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 1-115 FROM N.A.  
RX MEDLINE=86136024; PubMed=3004972;

RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,  
 RA Stunnenberg H., Bujard H.;  
 RT "Polymorphism of the precursor for the major surface antigens of  
 RT Plasmodium falciparum merozoites: studies at the genetic level";  
 RL EMBO J. 4:3823-3829(1985).  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR  
 CC (POTENTIAL).  
 CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42  
 CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF  
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X05624; CAA29112.1; -  
 DR PIR; A26868; A26868.  
 DR PIR; B25120; B25120.  
 DR InterPro; IPR000561; EGF-like.  
 DR Pfam; PF00008; EGF; 1.  
 KW Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;  
 KW Transmembrane; GPI-anchor.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.  
 FT CARBOHYD 110 110 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 536 536 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 607 607 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 802 802 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 899 899 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 919 919 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 965 965 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 991 991 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 1701 AA; 193768 MW; 3FC2EC59AF96EA98 CRC64;

Query Match 57.4%; Score 42.5; DB 1; Length 1701;  
 Best Local Similarity 60.0%; Pred. No. 11;  
 Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

Oy 1 QYIKANSKFI-GITE 14  
 I:::|||||I::I  
 Db 1001 QFVKSNSKVIITGLTE 1015

RESULT 6  
 MSP1\_PLAFC STANDARD; PRT; 1726 AA.  
 AC P04934;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)  
 DE (PMMSA) (P195).  
 GN MSP-1.  
 OS Plasmodium falciparum (isolate Camp / Malaysia).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5835;  
 RN [1]  
 RP SEQUENCE OF 1-1103 FROM N.A.  
 RX MEDLINE=86205236; PubMed=3517809;  
 RA Weber J.L., Leininger W.M., Lyon J.A.;  
 RT "Variation in the gene encoding a major merozoite surface antigen of  
 RT the human malaria parasite Plasmodium falciparum";  
 RL Nucleic Acids Res. 14:3311-3323(1986).

RN SEQUENCE OF 1104-1726 FROM N.A.  
 RP MEDLINE=88143999; PubMed=3278296;  
 RA Weber J.L., Sim B.K.L., Lyon J.A., Wolff R.;  
 RT "Merozoite surface protein sequence from the Camp strain of the human  
 RT malaria parasite Plasmodium falciparum";  
 RL Nucleic Acids Res. 16:1206-1206(1988).  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR  
 CC (POTENTIAL).  
 CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42  
 CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF  
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X03831; CAA27446.1; -  
 DR PIR; A23386; SAZQCM.  
 DR InterPro; IPR000561; EGF-like.  
 DR Pfam; PF00008; EGF; 1.  
 KW Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;  
 KW Transmembrane; GPI-anchor.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1.  
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 567 567 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 638 638 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 827 827 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 944 944 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 990 990 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 1726 AA; 196197 MW; DD8AD45FA352BCF3 CRC64;

Query Match 57.4%; Score 42.5; DB 1; Length 1726;  
 Best Local Similarity 60.0%; Pred. No. 11;  
 Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

Oy 1 QYIKANSKFI-GITE 14  
 I:::|||||I::I  
 Db 1026 QFVKSNSKVIITGLTE 1040

RESULT 7  
 MSP1\_PLAFC STANDARD; PRT; 1726 AA.  
 AC P50435;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)  
 DE (PMMSA) (GPI95).  
 GN MSP-1.  
 OS Plasmodium falciparum (isolate Palo Alto / Uganda).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=57270;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89005525; PubMed=3049134;  
 RA Chang S.P., Kramer K.J., Yamaga K.M., Kato A., Case S.E.,  
 RA Siddiqui W.A.;  
 RT "Plasmodium falciparum: gene structure and hydropathy profile of the

RT major merozoite surface antigen (gp195) of the Uganda-Palo Alto  
RT isolate.";  
RL Exp. Parasitol. 67:1-11(1988).  
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR  
CC (POTENTIAL).  
CC  
CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42  
CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF  
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; M37213; AAA29611.1; -  
DR InterPro: IPR000561; EGF-like.  
DR Pfam: PF00008; EGF; 1.  
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;  
KW Transmembrane; GPI-anchor.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1.  
FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 272 272 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 567 567 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 638 638 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 827 827 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 924 924 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 944 944 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 990 990 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT SEQUENCE 1726 AA; 196174 MW; 5B59CEEFA2F9A026 CRC64;  
SQ  
  
Query Match 57.4%; Score 42.5; DB 1; Length 1726;  
Best Local Similarity 60.0%; Pred. No. 11;  
Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;  
  
QY 1 QVVKANSKFI-GITE 14  
DB 1026 QVVKNSKVITGLTE 1040  
I:::|::| |::|  
  
RESULT 8  
ID PYRC\_SERMA STANDARD; PRT; 204 AA.  
AC Q9S3S1;  
DT 20-AUG-2001 (Rel. 40, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE DIHYDROOROTASE (EC 3.5.2.3) (DHQASE) (FRAGMENT).  
GN PYRC.  
OS Serratia marcescens.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Serratia.  
OX NCBI\_TaxID=615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SM6;  
RA Berkmen M., Benedik M.J.;  
RT "Dini inhibits transcription of Serratia marcescens nuclease.";  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- CATALYTIC ACTIVITY: DIHYDROOROTATE + H(2)O = N-CARBAMOYL-  
CC L-ASPARTATE.  
CC  
CC -1- COFACTOR: THIS ENZYME TIGHTLY BINDS ONE ZINC ATOM PER CHAIN WHICH  
CC IS REQUIRED FOR THE CATALYTIC MECHANISM. IT ALSO BINDS WEAKLY TO  
CC TWO OTHERS ZINCS WHICH ARE NOT ESSENTIAL FOR ACTIVITY (BY

CC SIMILARITY).  
CC -1- PATHWAY: THIRD STEP IN PYRIMIDINE BIOSYNTHESIS.  
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE DHQASE FAMILY. SUBFAMILY 1.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; AF175466; AAD50307.1; ALT INIT.  
DR InterPro: IPR002195; Dihydroorotase.  
DR PROSITE; PS00482; DIHYDROOROTASE\_1; PARTIAL.  
DR PROSITE; PS00483; DIHYDROOROTASE\_2; 1.  
KW Pyrimidine biosynthesis; Hydrolase; Zinc.  
FT NON\_TER 1  
SQ SEQUENCE 204 AA; 22899 MW; 9DB0B6C9B834B310 CRC64;  
  
Query Match 55.4%; Score 41; DB 1; Length 204;  
Best Local Similarity 46.2%; Pred. No. 2.3;  
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 QYIKANSKFIGIT 13  
DB 42 QYVQAGNRFLGAT 54  
I:::|::| |::|  
  
RESULT 9  
ID V120\_HSVSA STANDARD; PRT; 899 AA.  
AC Q01055;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE CAPSID ASSEMBLY PROTEIN 63.  
GN 63 OR PERFL.  
OS Herpesvirus saimiri (strain 11).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Gammaherpesvirinae; Rhadinovirus.  
OX NCBI\_TaxID=10383;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92333688; PubMed=1321287;  
RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Blesinger B.,  
RA Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,  
RA Honess R.W.;  
RT "Primary structure of the herpesvirus saimiri genome.";  
RL J. Virol. 66:5047-5058(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92330228; PubMed=1314457;  
RA Nicholas J., Cameron K.R., Coleman H., Newman C., Honess R.W.;  
RT "Analysis of nucleotide sequence of the rightmost 43 kbp of  
RT herpesvirus saimiri (HVS) L-DNA: general conservation of genetic  
RT organization between HVS and Epstein-Barr virus.";  
RL Virology 188:296-310(1992).  
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL37,  
CC EBV-1 23, EBV BOLFI, VZV 21, HVS-1 63, AND HCMV UL47.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; X64346; CAA45686.1; -  
DR EMBL; M86409; AAA46139.1; -

```
DR PIR; G36812; G36812.
KW Capsid assembly.
SQ SEQUENCE 899 AA; 103350 MW; F1429B3770A2885E CRC64;

Query Match 55.4%; Score 41; DB 1; Length 899;
Best Local Similarity 50.0%; Pred. No. 11;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 QYIRANSKFIGITE 14
III : I : I : I :
Db 124 QYITSNATFTGLSE 137

RESULT 10
ACPD_HAEIN STANDARD; PRT; 194 AA.
AC P43013;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ACYL CARRIER PROTEIN PHOSPHODIESTERASE (EC 3.1.4.14) (ACP
DE PHOSPHODIESTERASE).
GN ACPD OR HI1366.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OC Haemophilus.
OX NCBI_TaxID=727;
[1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=96186898; PubMed=8635745;
RA Chandler M.S., Smith R.A.;
RT "Characterization of the Haemophilus influenzae topA locus: DNA
RT topoisomerase I is required for genetic competence.";
RL Gene 169:25-31(1996).
[2]
RN RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fieschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -1- FUNCTION: CONVERTS HOLO-ACP TO APO-ACP BY HYDROLYTIC CLEAVAGE OF
CC THE PHOSPHOPANTHETHEINE RESIDUE FROM ACP (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: [ACYL-CARRIER PROTEIN] + H2O -> 4-
CC PHOSPHOPANTHETHEINE + APOPROTEIN.
CC -1- SIMILARITY: BELONGS TO THE ACPD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U20964; AAC43728.1; -.
CC DR EMBL; U32816; AAC23013.1; -.
CC DR TIGR; HI1366; -.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 194 AA; 21208 MW; A8AEC9D00829522C CRC64;

Query Match 54.1%; Score 40; DB 1; Length 601;
Best Local Similarity 46.7%; Pred. No. 11;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15
III : I : I : I : I :
Db 147 QYMKSIILGFIGITDV 161

Query Match 54.1%; Score 40; DB 1; Length 194;
Best Local Similarity 53.3%; Pred. No. 3;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15
III : I : I : I : I :
Db 147 QYMKSIILGFIGITDV 161

RESULT 11
PEFL_LACIC STANDARD; PRT; 601 AA.
ID PEFL_LACLC
AC P54124; P94880;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE OLIGOENDOPEPTIDASE F, PLASMID (EC 3.4.24.-).
DE PEPLI OR PEPPF.
GN Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OS Lactococcus lactis.
OG Plasmid pLP763.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1359;
[1]
RN RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=NCDO 763;
RX MEDLINE=95096044; PubMed=7798200;
RA Monnet V., Nardi M., Chopin A., Chopin M.-C., Gripon J.-C.;
RT "Biochemical and genetic characterization of PepF, an oligopeptidase
RT from Lactococcus lactis.";
RL J. Biol. Chem. 269:32070-32076(1994).
[2]
RN RP SEQUENCE FROM N.A.
RC STRAIN=NCDO 763;
RX MEDLINE=97352670; PubMed=9209029;
RA Nardi M., Renault P., Monnet V.;
RT "Duplication of the pepF gene and shuffling of DNA fragments on the
RT lactose plasmid of Lactococcus lactis.";
RL J. Bacteriol. 179:4164-4171(1997).
CC -1- FUNCTION: HYDROLYZES PEPTIDES CONTAINING BETWEEN 7 AND 17 AMINO
CC ACIDS WITH A RATHER WIDE SPECIFICITY.
CC -1- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE);
CC ALSO KNOWN AS THE THIMET OLIGOPEPTIDASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z32522; CAA83534.1; -.
CC DR EMBL; X95798; CAA68133.1; -.
CC DR MEROPS; M03.007; -.
CC DR InterPro; IPR001567; Peptidase_M3.
CC DR InterPro; IPR000130; Zn_MpPeptidse.
CC DR Pfam; PF01432; Peptidase_M3; 1.
CC DR PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
CC KW Hydrolase; Metalloprotease; Zinc; Plasmid.
CC METAL; 387 387 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT ACCT_SITE 388 388 BY SIMILARITY.
CC FT METAL 391 391 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT METAL 394 394 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT CONFLICT 518 518 F -> S (IN REF. 2).
CC SQ SEQUENCE 601 AA; 69674 MW; C8B5C519FFA2F787 CRC64;
```

```
QY 1 QYIKANSKFIGITEL 15
DB 284 RYIELRKKILGITDL 298

RESULT 12
YHJ9_YEAST
ID YHJ9_YEAST STANDARD; PRT; 644 AA.
AC P38694;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL ALDEHYDE-DEHYDROGENASE LIKE PROTEIN IN FIL1-VMA10
DE INTERGENIC REGION.
GN YHR039C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favell A., Fulton L., Gattung S., Giesel C., Kirsten J.,
RA Kucaba T., Hillier L., Jler M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macris C., Mardis E., Meneses S., Mouser L.,
RA Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII."
RL Science 265:2077-2082(1994).
CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U00062; AAB68915.1; -
DR PIR; S46746; S46746.
DR SGD; S0001081; YHR039C.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE-DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE-DEHYDR_GLU; 1.
KW Hypothetical protein; Oxidoreductase.
FT ACT_SITE 354 354 BY SIMILARITY.
FT ACT_SITE 389 389 BY SIMILARITY.
SQ SEQUENCE 644 AA; 71320 MW; 54DADDAEB2A16D4D CRC64;

Query Match 54.1%; Score 40; DB 1; Length 644;
Best Local Similarity 60.0%; Pred. No. 12;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15
DB 38 QIIQDNQKLIGITL 52

RESULT 13
KAD_MYCCA
ID KAD_MYCCA STANDARD; PRT; 213 AA.
AC P10251;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ADENYLATE KINASE (EC 2.7.4.3) (ATP-AMP TRANSPHOSPHORYLASE).
```

---

```
GN ADK.
OS Mycoplasma capricolum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Entomoplasmataceae.
OX NCBI_TaxID=2095;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 27343 / KID;
RX MEDLINE=88142549; PubMed=3481422;
RA Okkubo S., Muto A., Kawauchi Y., Yamao F., Osawa S.;
RT "The ribosomal protein gene cluster of Mycoplasma capricolum.";
RL Mol. Gen. Genet. 210:314-322(1987).
CC -1- FUNCTION: THIS SMALL UBIQUITOUS ENZYME IS ESSENTIAL FOR
CC MAINTENANCE AND CELL GROWTH.
CC -1- CATALYTIC ACTIVITY: ATP + AMP = ADP + ADP.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE ADENYLATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X06414; CAA29724.1; -
DR PIR; S02851; KIYMC.
DR HSSP; P27142; LZIO.
DR InterPro; IPR000850; Adenylate_kin.
DR Pfam; PF00406; adenylatekinase; 1.
DR PRINTS; PR00094; ADENYLTKINASE.
DR PRODOM; PD000657; Adenylate_kin; 1.
DR PROSITE; PS00113; ADENYLATE_KINASE; 1.
KW Transferase; Kinase; ATP-binding.
FT NP_BIND 7 15 ATP (BY SIMILARITY).
SQ SEQUENCE 213 AA; 24616 MW; FBCFA426B6F92E16 CRC64;

Query Match 52.7%; Score 39; DB 1; Length 213;
Best Local Similarity 72.7%; Pred. No. 5.6;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 YIKANSKFIGI 12
DB 183 YFRTNSKFIETI 193

RESULT 14
YD83_METJA
ID YD83_METJA STANDARD; PRT; 256 AA.
AC Q58778;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN MJ1383 PRECURSOR.
GN MJ1383.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
```

RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
jannaschii";  
RL Science 273:11058-1073(1996).  
CC -!- SIMILARITY: TO M.JANNASCHII MJ0761.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U67578; AAB99393.1; -  
DR TIGR; MJ1383; -  
DR InterPro; IPR001130; UPF00006.  
DR Pfam; PF01026; TatD.DNase; 1.  
KW Hypothetical protein; Signal; Complete proteome.  
FT SIGNAL 1 30 POTENTIAL.  
FT CHAIN 31 256 HYPOTHETICAL PROTEIN MJ1383.  
SQ SEQUENCE 256 AA; 29036 MW; 6D26F427EAB54675 CRC64;  
  
Query Match 51.4%; Score 38; DB 1; Length 256;  
Best Local Similarity 50.0%; Pred. No. 10;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
  
QY 2 YIKANSKFIGITEL 15  
II II: :|| I;  
Db 89 YINYSRVVGIGEI 102  
  
RESULT 15  
ID TRUB\_AQUAE STANDARD; PRT; 287 AA.  
AC O66922;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DE 20-AUG-2001 (Rel. 40, Last annotation update)  
DE TRNA PSEUDOURIDINE SYNTHASE B (EC 4.2.1.70) (TRNA PSEUDOURIDINE 55  
DE SYNTHASE) (PSI55 SYNTHASE) (PSEUDOURIDYLATE SYNTHASE) (URACIL  
DE HYDROLYASE).  
DE TRUB OR AQ.705.  
OS Aquifex aeolicus.  
OC Bacteria; Aquificales; Aquificaceae; Aquifex.  
OX NCBI\_TaxID=63363;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VF5;  
RX MEDLINE=98196666; PubMed=9537320;  
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
RA Graham D.E., Overbeek R., Shead M.A., Keller M., Aufay M., Huber R.,  
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
RT "The complete genome of the hyperthermophilic bacterium Aquifex  
RT aeolicus";  
RL Nature 392:353-358(1998).  
CC -!- FUNCTION: FORMATION OF PSEUDOURIDINE AT POSITION 55 IN THE  
CC PSI GC LOOP OF TRANSFER RNAs (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: URACIL + D-RIBOSE 5-PHOSPHATE = PSEUDOURIDINE  
CC 5'-PHOSPHATE + H(2)O.  
CC -!- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE000703; AAC06885.1; -  
DR InterPro; IPR002501; Trub\_N.

DR Pfam; PF01509; Trub\_N; 1.  
KW Lyase; tRNA processing; Complete proteome.  
SQ SEQUENCE 287 AA; 32259 MW; 2E58C1B7CEAD58AA CRC64;

Query Match 51.4%; Score 38; DB 1; Length 287;  
Best Local Similarity 69.2%; Pred. No. 12;  
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
  
QY 3 IKANSKFIGITEL 15  
II :||||| II  
Db 263 IYDSKFIGIGEL 275

Search completed: January 29, 2002, 11:16:20  
Job time: 149 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OH protein - protein search, using sw model

Run on: January 29, 2002, 11:15:53 ; Search time 62.77 Seconds  
(without alignments)  
51.266 Million cell updates/sec

Title: US-09-763-397a-25  
Perfect score: 108  
Sequence: 1 MKELVNVALVFVWYISYIYAD 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP-archaea:\*  
2: SP-bacteria:\*  
3: SP-fungi:\*  
4: SP-human:\*  
5: SP-invertebrate:\*  
6: SP-mammal:\*  
7: SP-mhc:\*  
8: SP-organella:\*  
9: SP-phage:\*  
10: SP-plant:\*  
11: SP-rodent:\*  
12: SP-virus:\*  
13: SP-vertebrate:\*  
14: SP-unclassified:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55	50.9	205	2	09XD01
2	54	50.0	205	2	09XD00
3	52	48.1	209	2	046325
4	52	43.1	209	2	09S4S8
5	52	48.1	209	2	09R416
6	50.5	46.8	352	2	09CG28
7	49	45.4	669	5	09VPG2
8	48	44.4	40	8	034216
9	47	43.5	65	2	099TK2
10	47	43.5	157	13	098882
11	47	43.5	162	13	091AK7
12	47	43.5	162	13	091AK6
13	47	42.6	162	13	091AK5
14	46	42.6	940	13	073635
15	45	41.7	137	11	09JK21
16	45	41.7	320	5	09XUK6
17	45	41.7	331	12	09YK45
18	45	41.7	379	2	087181
19	45	41.7	425	5	019125

20	45	41.7	446	2	09S115	09s115 streptococ
21	45	41.7	506	5	021420	021420 caenorhabd
22	45	41.7	681	5	09VE33	09ve33 drosophila
23	45	41.7	1260	5	09XZC5	09xzcs cryptospori
24	44.5	41.2	210	8	09G8P4	09g8p4 naegleria g
25	44	40.7	222	13	P70023	P70023 xenopus lae
26	44	40.7	321	11	09QXFS	09qxfs mesocricetu
27	44	40.7	373	5	021149	021149 caenorhabd
28	44	40.7	377	2	034161	034161 salmone
29	44	40.7	492	5	027505	027505 caenorhabd
30	44	40.7	495	2	09X4D3	09x4d3 streptococ
31	44	40.7	580	5	016920	016920 caenorhabd
32	44	40.7	687	9	09MC93	09mc93 bacterioph
33	43.5	40.3	576	3	09Y7S4	09y7s4 schizosacch
34	43.5	40.3	580	3	059698	059698 schizosacch
35	43	39.8	94	10	09LYP8	09lyp8 arabidopsis
36	43	39.8	217	5	017717	017717 caenorhabd
37	43	39.8	254	2	09LBY3	09lby3 shewanella
38	43	39.8	269	10	039647	039647 cucurbita m
39	43	39.8	300	2	09KTI2	09kti2 vibrio chol
40	43	39.8	338	5	045300	045300 caenorhabd
41	43	39.8	372	3	042944	042944 schizosacch
42	43	39.8	379	13	09DG05	09dg05 gallus gall
43	43	39.8	424	13	09DG06	09dg06 gallus gall
44	43	39.8	442	1	058688	058688 pyrococcus
45	43	39.8	442	2	09HTR7	09htr7 pseudomonas

## ALIGNMENTS

RESULT 1	PRELIMINARY:	PRT:	205 AA.
ID 09XD01			
AC 09XD01			
DT 01-NOV-1999 (TREMBlrel. 12, Created)			
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)			
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)			
DE MAJOR ANTIGENIC PROTEIN 2 HOMOLOG.			
OS Ehrlichia canis.			
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;			
OC Rickettsiaceae; Ehrlichiae; Ehrlichia.			
OX NCBI_Taxid=944;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=OKLAHOMA;			
RA Bowie M.V., Reddy G.R., Semu S.M., Mahan S.M., Barbet A.F.;			
RT "Potential Value of Major Antigenic Protein 2 for serological			
RT Diagnosis of Heartwater and Related Ehrlichial Infections.";			
RL Clin. Diagn. Lab. Immunol. 0:0-0(1999).			
DR EMBL: AF17730; AAD40619.1; --			
DR InterPro: IPR003782; SCOL_Senc.			
DR Pfam: PF02630; SCOL_Senc; 1.			
SQ SEQUENCE 205 AA; 23169 MW; 756B50A3304C13CB CRC64;			
Query Match	50.9%	Score 55;	DB 2; Length 205;
Best Local Similarity	40.0%	Pred. No. 3.4;	
Matches 8; Conservative	7;	Mismatches 5;	Indels 0;
		Gaps 0;	
QY 1 MKELVNVALVFVWYISYIY 20			
DB 4 IKFILNVCLFAIRFLGYSY 23			
RESULT 2			
ID 09XD00			
AC 09XD00	PRELIMINARY:	PRT:	205 AA.
DT 01-NOV-1999 (TREMBlrel. 12, Created)			
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)			
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)			
DE MAJOR ANTIGENIC PROTEIN 2 HOMOLOG.			

OS Ehrlichia chaffeensis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Ehrlichiae; Ehrlichia.  
 OX NCBI\_TaxID=945;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ARKANSAS;  
 RA Bowie M.V., Reddy G.R., Semu S.M., Mahan S.M., Barbet A.F.;  
 RT "Potential Value of Major Antigenic Protein 2 for Serological  
 RT Diagnosis of Heartwater and Related Ehrlichial Infections."  
 RL Clin. Diagn. Lab. Immunol. 0:0-0(1999).  
 DR EMBL: AF117731; AAD40620.1;  
 DR InterPro: IPR003782; SCOL\_Senc.  
 DR Pfam: PF02630; SCOL\_Senc; 1.  
 SK  
 SQ SEQUENCE 205 AA; 23142 MW; 561DC264B6EC9736 CRC64;

Query Match 50.0%; Score 54; DB 2; Length 209;  
 Best Local Similarity 35.0%; Pred. No. 4.6;  
 Matches 7; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

OY 1 MKFLVNAVLFVNVYISYIY 20  
 DB 4 IKFLINCLLFAIFLGYSY 23

RESULT 3  
 ID 046325 PRELIMINARY; PRT; 209 AA.  
 AC 046325;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE IMMUNODOMINANT PROTEIN PRECURSOR.  
 OS Cowdria ruminantium.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Ehrlichiae; Cowdria.  
 OX NCBI\_TaxID=779;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=95005467; PubMed=7921263;  
 RA Mahan S.M., McQuire T.C., Semu S.M., Bowie M.V., Jongejan F.,  
 RA Rurangirwa F.R., Barbet A.F.;  
 RT "Molecular cloning of a gene encoding the immunogenic 21 kDa protein  
 RT of Cowdria ruminantium."  
 RL Microbiology 140:2135-2142(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HIGHWAY;  
 RA Bowie M.V., Reddy G.R., Semu S.M., Mahan S.M., Barbet A.F.;  
 RT "Potential Value of Major Antigenic Protein 2 for Serological  
 RT Diagnosis of Heartwater and Related Ehrlichial Infections."  
 RL Clin. Diagn. Lab. Immunol. 0:0-0(1999).  
 DR EMBL: L07383; AAS0280.1;  
 DR EMBL: AF117727; AAD40616.1;  
 DR InterPro: IPR003782; SCOL\_Senc.  
 DR Pfam: PF02630; SCOL\_Senc; 1.  
 DR Signal.  
 KM  
 FT SIGNAL  
 FT CHAIN 1 19 POTENTIAL.  
 FT CHAIN 20 209 IMMUNODOMINANT PROTEIN.  
 SQ SEQUENCE 209 AA; 23608 MW; 45D107A87F0B9345 CRC64;

Query Match 48.1%; Score 52; DB 2; Length 209;  
 Best Local Similarity 35.0%; Pred. No. 8.7;  
 Matches 7; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

OY 1 MKFLVNAVLFVNVYISYIY 20  
 DB 8 IKFLINCLLFAIFLGYSY 27

RESULT 4

O9S4S8  
 ID O9S4S8 PRELIMINARY; PRT; 209 AA.  
 AC O9S4S8;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE MAJOR ANTIGENIC PROTEIN 2.  
 OS Cowdria ruminantium.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Ehrlichiae; Cowdria.  
 OX NCBI\_TaxID=779;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ANTIGUA;  
 RA Bowie M.V., Reddy G.R., Semu S.M., Mahan S.M., Barbet A.F.;  
 RT "Potential Value of Major Antigenic Protein 2 for Serological  
 RT Diagnosis of Heartwater and Related Ehrlichial Infections."  
 RL Clin. Diagn. Lab. Immunol. 0:0-0(1999).  
 DR EMBL: AF117726; AAD40615.1;  
 DR InterPro: IPR003782; SCOL\_Senc.  
 DR Pfam: PF02630; SCOL\_Senc; 1.  
 SK  
 SQ SEQUENCE 209 AA; 23504 MW; 8FF9A535E1737569 CRC64;

Query Match 48.1%; Score 52; DB 2; Length 209;  
 Best Local Similarity 35.0%; Pred. No. 8.7;  
 Matches 7; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

OY 1 MKFLVNAVLFVNVYISYIY 20  
 DB 8 IKFLINCLLFAIFLGYSY 27

RESULT 5  
 ID O9R416 PRELIMINARY; PRT; 209 AA.  
 AC O9R416;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE MAJOR ANTIGENIC PROTEIN 2.  
 OS Cowdria ruminantium.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Ehrlichiae; Cowdria.  
 OX NCBI\_TaxID=779;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=UM BANEIN, AND PALM RIVER;  
 RA Bowie M.V., Reddy G.R., Semu S.M., Mahan S.M., Barbet A.F.;  
 RT "Potential Value of Major Antigenic Protein 2 for Serological  
 RT Diagnosis of Heartwater and Related Ehrlichial Infections."  
 RL Clin. Diagn. Lab. Immunol. 0:0-0(1999).  
 DR EMBL: AF117729; AAD40617.1;  
 DR EMBL: AF117728; AAD40617.1;  
 DR InterPro: IPR003782; SCOL\_Senc.  
 DR Pfam: PF02630; SCOL\_Senc; 1.  
 SK  
 SQ SEQUENCE 209 AA; 23562 MW; 1C3699C5DE718ABB CRC64;

Query Match 48.1%; Score 52; DB 2; Length 209;  
 Best Local Similarity 35.0%; Pred. No. 8.7;  
 Matches 7; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

OY 1 MKFLVNAVLFVNVYISYIY 20  
 DB 8 IKFLINCLLFAIFLGYSY 27

RESULT 6  
 ID O9CGZ8 PRELIMINARY; PRT; 352 AA.  
 AC O9CGZ8;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)

Query Match	46.8%;	Score 50.5;	DB 2;	Length 35;
Best Local Similarity	43.5%;	Pred. No. 22;		
Matches	10;	Conservative	6;	Mismatches 4; Indels 3; Gaps 1
Qy	2	KFLVNALVPMV--YISYTA	21	
	1	11::111::1::1:11		
Db	10	KILVSIILVFALIGSVYIYVA	32	

RESULT	7	
09VPG2		
ID	09VPG2	PRELIMINARY; PRT; 669 AA.
AC	09VPG2;	
DT	01-MAY-2000 (TREMBLrel. 13, Created)	
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)	
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)	
DE	CG13248 PROTEIN.	
GN	CG13248	
OS	<i>Drosophila melanogaster</i> (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;	
OC	Ephydroidea; Drosophilidae; <i>Drosophila</i> .	
OX	NCBI_TaxID=7227;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-BERKELEY;	
RX	MEDLINE-20196006; PubMed-10731132;	
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,	
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	
RA	Sutten G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,	
RA	Brondon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Miller B.D.,	
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,	
RA	Abdl J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,	
RA	Ballew R.M., Basu A., Bakendole J., Bayraktaroglu L., Beasley E.M.,	
RA	Berson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,	
RA	Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,	
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,	
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,	
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,	
RA	Dubin K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,	
RA	Foster K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,	
RA	Glodok A., Gong F., Gorrell J.H., Gu Z., Gunn P., Harris M.,	
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,	
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,	
RA	Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,	
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,	
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lai X.,	
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,	
RA	Merkulov G., Mishina N.V., Moberly C., Morris J., Moshireli A.,	
RA	Mount S.M., Moy M., Murphy B., Murphy L., Mutny D.M., Nelson D.L.,	

		45.4%	Score	49;	DB	5;	Length	669;	
Query Match		Best Local Similarity	42.1%;	Pred.	No.	60;			
Matches	8;	Conservative	7;	Mismatches	4;	Indels	0;	Gaps	0
Qy	4	LNNVALYENVIISITYAD	22	.					
	: : :	: : : :		-					
Db	202	LNNIAVAVLVISGFEYAD	220						

ID	034216	PRELIMINARY;	PRT;	40 AA.
AC	034216:			
DT	01-NOV-1996 (TRMBLrel. 01, Created)			
DT	01-NOV-1996 (TRMBLrel. 01, Last sequence update)			
DT	01-NOV-1998 (TRMBLrel. 08, Last annotation update)			
DE	NADH DEHYDROGENASE SUBUNIT 2 (FRAGMENT).			
GN	ND2.			
OS	Candida parapsilosis (Yeast).			
OC	Mitochondrion.			
CC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
CC	Saccharomycetales; mitosporic Saccharomycetales; Candida.			
OX	NCBI_TaxId=5480;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=SR2;			
RX	MEDLINE=94364940; PubMed=7521869;			
RA	Nosek J., Fukuhara H.;			
RT	"NADH dehydrogenase subunit genes in the mitochondrial DNA of yeasts."			
RL	J. Bacteriol. 176:5622-5630(1994).			
DR	EMBL: X75676; CAA53373.1; -.			
KM	Mitochondrion.			
FT	NON PER			
SO	SEQUENCE	1	1	
		40 AA;	4665 MW;	7A7CEP09FE1BE376 CRC64;

Query Match	44.4%;	Score 48;	DB 8;	Length 40;
Best Local Similarity	35.0%;	Pred. No. 7.4;		
Matches	7;	Conservative	7;	Mismatches 6;
			Indels 0;	Gaps 0;
Qy	1	MKFLVNAVLEPMVYISYIY	20	
	:		:	:
Db	10	LAFVLSLLVFFIFYFYIY	29	
RESULT	9			
Q099TK2		PRELIMINARY;	PRT:	65 AA.
ID	Q099TK2			
AC	Q099TK2:			
DT	01-JUN-2001	(TREMBlrel, 17, Created)		
DT	01-JUN-2001	(TREMBlrel, 17, Last sequence update)		

01-JUN-2001 (TREMblrel. 17, last annotation update)  
 HYPOTHETICAL PROTEIN SA1477.  
 SA1477  
 GN Staphylococcus aureus subsp. aureus N315.  
 OS Staphylococcus aureus  
 OC Bacillus/Staphylococcus group; Staphylococcus.  
 DE Bacillus/Staphylococcus group; Staphylococcus.  
 NCBI\_TaxID=158679;  
 [1]  
 RP  
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
 Cui L., Oguchi A., Aoki K.I., Nagai Y., Ito T., Kanamori M.,  
 Takamatsu H., Matsumura A., Murakami H., Hoshiyama A., Mizutani-U Y.,  
 Takahashi N.K., Sawano T., Inoue R.I., Kato C., Sekimizu K.,  
 Yamashita H., Kohda S., Goto S., Terauchi J., Kashiwa M.,  
 Yamashita K., Oshima K., Ruvinsky A., Yoshino C., Shiba T., Hattori M.,  
 Ogasawara N., Hayashi H., Hirumatsu K.;  
 "Whole genome sequencing of methicillin-resistant Staphylococcus  
 aureus.";  
 RT Lancet 357:1225-1240(2001).  
 DR EMBL: AP003134; BAB2743.1;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 65 AA; 7317 MW; D875C82A08C70E8 CRC64;

Query Match 43.5%; Score 47; DB 13; Length 65;  
 Best Local Similarity 38.9%; Pred. No. 15;  
 Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;  
 OY 1 LKLVNVALVPMVYISY 18  
 DB 24 LKLVNVALVPMVYISY 41

RESULT 10  
 ID O98882 PRELIMINARY; PRT; 157 AA.  
 AC O98882;  
 DT 01-FEB-1997 (TREMblrel. 02, Created)  
 DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE ODOURANT RECEPTOR 1 (FRAGMENT).  
 GN ZORI.  
 OS Brachydanio rerio (zebrafish) (zebra danio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
 OC Cypriniformes; Cyprinidae; Rasbora; Danio.  
 NCBI\_TaxID=7955;  
 RN [1]  
 RP  
 RA "SEQUENCE FROM N.A.  
 RA MEDLINE-97025163; PubMed-8917589;  
 RA Weich F., Nadler W., Korsching S.  
 RT "Nested expression domains for odorant receptors in zebrafish  
 olfactory epithelium.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 93:13321-13326(1996).  
 DR EMBL: U72683; AAB38866.1;  
 DR ZFIN: ZDB-GENE-990415-190; Zcrl.  
 DR InterPro: IPR000276; GPCR\_Rhodopsn.  
 DR Pfam: PF00001; 2tm\_1; 1.  
 DR PROSITE: PS50262; G\_PROTEIN\_RECPT\_FL\_2; 1.  
 FT NON\_TER 157  
 FT NON\_TER 157  
 SQ SEQUENCE 157 AA; 17656 MW; CE5738C349E526C4 CRC64;

Query Match 43.5%; Score 47; DB 13; Length 157;  
 Best Local Similarity 41.2%; Pred. No. 32;  
 Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;  
 OY 4 LKLVNVALVPMVYISY 20  
 DB 77 LKLVNVALVPMVYISY 93

RESULT 11  
 ID O91AK7 PRELIMINARY; PRT; 162 AA.  
 AC O91AK7;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE OLFACTORY RECEPTOR (FRAGMENT).  
 GN DREL.  
 OS Brachydanio rerio (zebrafish) (zebra danio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
 OC Cypriniformes; Cyprinidae; Rasbora; Danio.  
 NCBI\_TaxID=7955;  
 RN [1]  
 RP  
 RA "SEQUENCE FROM N.A.  
 RA MEDLINE-20183981; PubMed-10706615;  
 RA Rouquier S., Blancher A., Giorgi D.;  
 RT "The olfactory receptor gene repertoire in primates and mouse:  
 RT Evidence for reduction of the functional fraction in primates.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 97:2870-2874(2000).  
 DR EMBL: AF179846; AAF40409.1;  
 DR InterPro: IPR000276; GPCR\_Rhodopsn.  
 DR PROSITE: PS50262; G\_PROTEIN\_RECPT\_FL\_2; 1.  
 FT NON\_TER 162  
 FT NON\_TER 162  
 SQ SEQUENCE 162 AA; 18156 MW; FA77B0A46EFCB7A9 CRC64;

Query Match 43.5%; Score 47; DB 13; Length 162;  
 Best Local Similarity 41.2%; Pred. No. 33;  
 Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;  
 OY 4 LKLVNVALVPMVYISY 20  
 DB 81 LKLVNVALVPMVYISY 97

RESULT 12  
 ID O91AK6 PRELIMINARY; PRT; 162 AA.  
 AC O91AK6;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE OLFACTORY RECEPTOR (FRAGMENT).  
 GN DREL.  
 OS Brachydanio rerio (zebrafish) (zebra danio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
 OC Cypriniformes; Cyprinidae; Rasbora; Danio.  
 NCBI\_TaxID=7955;  
 RN [1]  
 RP  
 RA "SEQUENCE FROM N.A.  
 RA MEDLINE-20183981; PubMed-10706615;  
 RA Rouquier S., Blancher A., Giorgi D.;  
 RT "The olfactory receptor gene repertoire in primates and mouse:  
 RT Evidence for reduction of the functional fraction in primates.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 97:2870-2874(2000).  
 DR EMBL: AF179846; AAF40409.1;  
 DR InterPro: IPR000276; GPCR\_Rhodopsn.  
 DR Pfam: PF00001; 2tm\_1; 1.  
 DR PROSITE: PS50262; G\_PROTEIN\_RECPT\_FL\_2; 1.  
 FT NON\_TER 162  
 FT NON\_TER 162  
 SQ SEQUENCE 162 AA; 18122 MW; 7877B0A46EFCB7A9 CRC64;

Query Match 43.5%; Score 47; DB 13; Length 162;  
 Best Local Similarity 41.2%; Pred. No. 33;  
 Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 4 LVNVALVFMVYISYIY 20  
1 : 1 : 1 : 1 : 1 :  
DB 81 LIGLLTFLILISYIY 97

## RESULT 13

09JAK5 PRELIMINARY; PRT; 162 AA.  
AC 09JAK5.  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE OLFACTORY RECEPTOR (FRAGMENT).  
GN DREG.  
OS Brachydanio rerio (zebrafish) (zebra danio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
OC Cypriniformes; Cyprinidae; Rasbora; Danio.  
NCBI\_TaxID=7955;  
RN 11  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20183981; PubMed=10706615;  
RA Rouquier S., Blancher A., Giorgi D.;  
RT "The olfactory receptor gene repertoire in primates and mouse:  
RT Evidence for reduction of the functional fraction in primates.";  
RL Proc. Natl. Acad. Sci. U.S.A. 97:2870-2874(2000).  
DR EMBL: AF179846; AAF0411.1;  
DR InterPro: IPR000276; GPCR\_Rhodopsin.  
DR Pfam: PF00001; 7tm.1; 1.  
DR PROSITE: PS50262; G\_PROTEIN\_RECP\_FL2; 1.  
KW Receptor.  
FT NON\_TER 1  
FT NON\_TER 162  
SQ SEQUENCE 162 AA; 18130 MW; E1FD981DD448C39 CRC64;

Query Match 43.5%; Score 47; DB 13; Length 162;  
Best Local Similarity 41.2%; Pred. No. 33;  
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 4 LVNVALVFMVYISYIY 20  
1 : 1 : 1 : 1 : 1 :  
DB 81 LIGLLTFLILISYIY 97

## RESULT 14

073635 PRELIMINARY; PRT; 940 AA.  
AC 073635;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE CALCIUM2+ SENSING RECEPTOR.  
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;  
OC Tetraodontidae; Takifugu.  
NCBI\_TaxID=31033;  
RN 11  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98226788; PubMed=9560249;  
RA Saito T., Saito Y., Yamamoto J., Nozaki Y., Tomura K., Hazama M.,  
RA Nakanishi S., Brenner S.;  
RT "Putative pheromone receptors related to the Ca2+-sensing receptor in  
RT Fugu.";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:5178-5181(1998).  
DR EMBL: AB008857; BAA2612.1;  
DR InterPro: IPR000337; GPCR\_Mgr.  
DR InterPro: IPR000631; RasGEF.  
DR InterPro: IPR001828; ANF\_receptor.

DR Pfam: PF00003; 7tm.3; 1.  
DR Pfam: PF01094; ANF\_receptor; 1.  
DR PRINTS: PR00248; GPCR\_MGR.  
DR PROSITE: PS00979; G\_PROTEIN\_RECP\_F3\_1; 1.  
DR PROSITE: PS50259; G\_PROTEIN\_RECP\_F3\_4; 1.  
SQ SEQUENCE 940 AA; 105813 MW; 06DAB7803B6878B3 CRC64;

## Query Match

Best Local Similarity 42.6%; Score 46; DB 13; Length 940;  
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

OY 2 KFLVVALVFMVYISYIY 21  
1 : 1 : 1 : 1 : 1 :  
DB 790 KFLVVALVFMVYISYIY 809

## RESULT 15

09JKZ1 PRELIMINARY; PRT; 137 AA.  
AC 09JKZ1.  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE CALCIUM-SENSING RECEPTOR (FRAGMENT).  
GN CASR.  
OS Meriones unguiculatus (Mongolian jird).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Gerbillinae;  
OC Meriones.  
NCBI\_TaxID=10047;  
RN 11  
RP SEQUENCE FROM N.A.  
RX TISSUE=SPIRAL MODIOLAR ARTERY, KIDNEY;  
RA Womberger K., Scofield M.A., Mangmann P.;  
RT "Evidence for a calcium sensing receptor in the vascular smooth muscle  
RT cells of the spiral modiolar artery.";  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF221064; AAF37826.1;  
DR InterPro: IPR000337; GPCR\_Mgr.  
DR PROSITE: PS00981; G\_PROTEIN\_RECP\_F3\_3; 1.  
DR PROSITE: PS50259; G\_PROTEIN\_RECP\_F3\_4; 1.  
KW Receptor.  
FT NON\_TER 1  
FT NON\_TER 137  
SQ SEQUENCE 137 AA; 15075 MW; 827656E1DC3306AC CRC64;

Query Match 41.7%; Score 45; DB 11; Length 137;  
Best Local Similarity 45.0%; Pred. No. 54;  
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

OY 2 KFLVVALVFMVYISYIY 21  
1 : 1 : 1 : 1 : 1 :  
DB 15 KFLVVALVFMVYISYIY 34

Search completed: January 29, 2002, 11:15:55  
Job time: 215 sec

Mon Feb 4 15:23:47 2002

us-09-763-397a-25.rsp